United States
Environmental Protection
Agency

EPA/600/R-07/106
July 2007

Transit Bus Load-Based Modal Emission Rate Model Development

Chapter 2 examines the diesel fuel combustion process and its relationship to diesel engine emissions formation.

Chapter 3 overviews the existing heavy-duty vehicle emission models and presents the proposed heavy-duty diesel vehicle modal emission model (HDDV-MEM).

Chapter 4 provides an overview of the emission rate testing databases provided by U.S. EPA, the quality assurance and quality control (QA/QC) procedures to review the validity of the data, and the methods used to post-process these databases to correct data deficiencies.

In Chapter 5, the various statistical models considered for data analysis are discussed.

Chapter 6 selects the database used to develop the conceptual model and discusses the influence of explanatory variables on emissions.

Chapter 7 covers sensitivity tests of driving mode definitions and outlines the potential impacts on derived models.

Chapter 8 elaborates the different emission models developed for idle, deceleration, acceleration and cruise driving modes.

EPA/600/R-07/106
July 2007

TRANSIT BUS LOAD-BASED MODAL EMISSION RATE MODEL DEVELOPMENT

by

Chunxia Feng
Randall Guensler
Michael Rodgers

School of Civil and Environmental Engineering
Georgia Institute of Technology
Atlanta, GA

Contract No: EP-05C-000033

EPA Project Officer: Sue Kimbrough
U.S. Environmental Protection Agency
National Risk Management Research Laboratory
Air Pollution Prevention and Control Laboratory
Research Triangle Park, NC 27711

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

ABSTRACT

Heavy-duty diesel vehicle (HDDVs) operations are a major source of oxides of nitrogen (NOx) and particulate matter (PM) emissions in metropolitan areas nationwide. Although HDDVs constitute a small portion of the onroad fleet, they typically contribute more than 45% of NOx and 75% of PM onroad mobile source emissions (U.S. EPA 2003). HDDV emissions are a large source of global greenhouse gas and toxic air containment emissions. Over the last several decades, both government and private industry have made extensive efforts to regulate and control mobile source emissions. The relative importance of emissions from HDDVs has increased significantly because today's gasoline powered vehicles are more than 95% cleaner than vehicles in 1968.

In current regional and microscale modeling conducted in every state except California, HDDV emissions rates are taken from the U.S. Environmental Protection Agency's (EPA's) MOBILE 6.2 model (U.S. EPA 2001a). The U.S. Environmental Protection Agency (U.S. EPA) is currently developing a new set of modeling tools for the estimation of emissions produced by onroad and off-road mobile sources. The new Multi-scale mOtor Vehicle & equipment Emission System, known as MOVES (U.S. EPA 2001a), is a modeling system designed to better predict emissions from onroad operations.

The major effort of this research is to develop a new heavy-duty vehicle load-based modal emission rate model that overcomes some of the limitations of existing models and emission rates prediction methods. This model is part of the proposed Heavy-Duty Diesel Vehicle Modal Emission Modeling (HDDV-MEM) which was developed by Georgia Institute of Technology (Guensler, et al. 2006). HDDV-MEM differs from other proposed HDDV modal models (Barth, et al. 2004; Frey, et al. 2002; Nam 2003) in that the modeling framework first predicts second-by-second engine power demand as a function of vehicle operating conditions and then applies brake-specific emission rates to these activity predictions.

FOREWORD

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The National Risk Management Research Laboratory (NRMRL) is the agency's center for investigation of technological and management approaches for preventing and reducing risks from pollution that threaten human health and the environment. The focus of the laboratory's research program is on methods and their cost-effectiveness for prevention and control of pollution to air, land, water, and subsurface resources; protection of water quality in public water systems; remediation of contaminated sites, sediments, and ground water; prevention and control of indoor air pollution; and restoration of ecosystems. NRMRL collaborates with both public and private sector partners to foster technologies that reduce the cost of compliance and to anticipate emerging problems. NRMRL's research provides solutions to environmental problems by: developing and promoting technologies that protect and improve the environment; advancing scientific and engineering information to support regulatory and policy decisions; and providing the technical support and information transfer to ensure implementation of environmental regulations and strategies at the national, state, and community levels.

This publication has been produced as part of the laboratory's strategic long-term research plan. It is published and made available by EPA's Office of Research and Development to assist the user community and to link researchers with their clients.

Sally Gutierrez, Director
National Risk Management Research Laboratory

EPA REVIEW NOTICE

This report has been peer and administratively reviewed by the U.S. Environmental Protection Agency and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ACRONYMS

%	percent
AADT	annual average daily traffic
AATA	Ann Arbor Transit Authority
Acc	acceleration
ANOVA	analysis of variance
APPCD	Air Pollution Prevention and Control Division
bhp	brake horsepower
BSFC	brake specific fuel consumption
C	Celsius
CARB	California Air Resources Board
CART	classification and regression testing
CE-CERT	College of Engineering - Center for Environmental Research and Technology
CMEM	Comprehensive Modal Emissions Model
CO	carbon monoxide
deg	degree
df	degrees of freedom
DPS	drag power surrogate
DVD	digital video disc
ECM	electronic control module
EMFAC	CARB's mobile source emission factor model
E(MS)	expected mean square
EPA	Environmental Protection Agency
F	Fahrenheit
FHWA	Federal Highway Administration
FR	Federal Register
FTP	Federal Test Procedure
g/bhp-hr	grams per brake-horsepower-hour
g/h	grams per hour
g/s	grams per second
GIS	geographic information system
GPS	global positioning system
GVWR	gross vehicle weight rating
HC	hydrocarbon
HDD	heavy-duty diesel
HDDV	heavy-duty diesel vehicle

HDDV-MEM	Heavy-Duty Diesel Vehicle-Modal Emission Model
HDV	heavy-duty vehicle
HDV8B	heavy-duty vehicle 8B
HDV-UDDS	heavy-duty vehicle urban dynamometer driving schedule
Hg	mercury
HHDDE	heavy-heavy duty diesel engine
HTBR	hierarchical tree-based regression
Hz	hertz
IC	internal combustion
IPS	inertial power surrogate
kPa	kilopascal
K/S	Kolmogorov-Smirnov
LAFY	Los Angeles freeway
LANF	Los Angeles non-freeway
lb	pound
lb-ft	pound-feet
LDV	light-duty vehicle
LHDDE	light-heavy duty diesel engine
MARTA	Metropolitan Atlanta Rapid Transit Authority
MDPV	medium duty passenger vehicle
MEASURE	Mobile Emissions Assessment System for Urban and Regional Scale Emissions
MM	method of moments
mg/m^3	milligrams per cubic meter
MHDDE	medium-heavy duty diesel engine
MOBILE	EPA's mobile source emission rate model
MOBILE6	EPA's mobile source emission rate model
MOVES	Motor Vehicle Emission Simulator
MPE	mean prediction error
mpg	miles per gallon
mph	miles per hour
mph/s	miles per hour per second
MS	mean square
N2	nitrogen
NAAQS	National Ambient Air Quality Standards
NCSU	North Carolina State University
NGM	EPA's Next Generation Model (mobile sources)
NIPER	National Institute for Petroleum and Energy Research
NIST	National Institute of Standards and Technology
NO	nitrogen oxide
NO_2	nitrogen dioxide
NONROAD	EPA's emission rate model for non-road sources

NO_X	nitrogen oxides
NRMRL	National Risk Management Research Laboratory
NYNF	New York non-freeway
O_2	oxygen
O_3	ozone
ODEC	Onroad Diesel Emissions Characterization
OLS	ordinary least squares
OTAQ	Office of Transportation and Air Quality
Pb	lead
PCV	positive crankcase ventilation
PERE	Physical Emission Rate Estimator
PM	particulate matter
PM_{10}	particulate matter ≤10 microns
$PM_{2.5}$	particulate matter ≤ 2.5 microns
ppmv	parts per million by volume
QA/QC	quality assurance/quality control
QQ	quantile-quantile
RARE	Regional Applied Research Effort
RMSE	root mean square error
RPM	revolutions per minute
SCFM	standard cubic feet per minute
SO_2	sulfur dioxide
SS	sum of squares
SSE	sum of squares due to errors
SSTO	total sum of squares
SUV	sport utility vehicle
TB-EPDS	transit bus engine power demand simulator
TIUS	Truck Inventory and Use Survey
TRB	Transportation Research Board
UCR-CERT	University of California Riverside – Center for Environmental Research and Technology
UDDS	urban dynamometer driving schedule
U.S. EPA	U.S. Environmental Protection Agency
UV	ultraviolet
VIF	variance inflation factor
VMT	vehicle miles traveled
VOCs	volatile organic compounds
VSP	vehicle specific power
$\mu g/m^3$	micrograms per cubic meter
µm	micron

SUMMARY

Heavy-duty diesel vehicle (HDDV) operations are a major source of pollutant emissions in major metropolitan areas. Accurate estimation of heavy-duty diesel vehicle emissions is essential in air quality planning efforts because highway and non-road heavy-duty diesel emissions account for a significant fraction of the oxides of nitrogen (NO_x) and particulate matter (PM) emissions inventories. MOBILE6 (U.S. EPA 2002a), EPA's mobile source emission rate model, uses an "average trip-based" approach to modeling as opposed to a more fundamental and robust modal modeling approach.

The major effort of this research is to develop a new heavy-duty vehicle load-based modal emission rate model that overcomes some of the limitations of existing models and emission rates prediction methods. This model is part of the proposed Heavy-Duty Diesel Vehicle Modal Emission Modeling (HDDV-MEM) which was developed by Georgia Institute of Technology. HDDV-MEM first predicts second-by-second engine power demand as a function of vehicle operating conditions and then applies brake-specific emission rates to these activity predictions.

To provide better estimates of microscale level emissions, this modeling approach is designed to predict second-by-second emissions from on-road vehicle operations. This research statistically analyzes the database provided by EPA and yields a model for prediction of emissions at a microscale level based on engine power demand and driving mode. Research results demonstrate the importance of including the influence of engine power demand vis-à-vis emissions and simulating engine power in real world applications. The modeling approach provides a significant improvement in HDDV emissions modeling compared to the current average speed cycle-based emissions models.

This page left blank deliberately.

CHAPTER 1

1. INTRODUCTION

1.1 Emissions from Heavy-Duty Diesel Vehicles

Heavy-duty diesel vehicles (HDDVs) operations are a major source of oxides of nitrogen (NO_x) and particulate matter (PM) emissions in metropolitan areas nationwide. Although HDDVs constitute a small portion of the on-road fleet, they typically contribute more than 45% of NO_x and 75% of PM on-road mobile source emissions (U.S. EPA 2003). HDDV emissions are a large source of global greenhouse gas and toxic air contaminant emissions. According to Environmental Defense Report in 2002, NO_x causes many environmental problems including acid rain, haze, global warming and nutrient overloading leading to water quality degradation (CEDF 2002). HDDV emissions are also harmful to human health and the environment (SCAQMD 2000). Groundbreaking long-term studies of children's health conducted in California have demonstrated that particle pollution may significantly reduce lung function growth in children (Avol 2001, Gauderman 2002, Peters 1999). Previous studies have stressed the significance of emissions from HDVs, in urban non-attainment areas especially for ozone (for which nitrogen oxides are a precursor) and $PM_{2.5}$ (Gautam and Clark 2003, Lloyd and Cackette 2001).

Over the last several decades, both government and private industry have made extensive efforts to regulate and control mobile source emissions. In 1961, the first automotive emissions control technology in the nation, Positive Crankcase Ventilation (PCV), was mandated by the California Motor Vehicle State Bureau of Air Sanitation to control hydrocarbon crankcase emissions, and PCV Requirement went into effect on domestic passenger vehicles for sale in California in 1963 (CARB 2004). At the same time, first Federal Clean Air Act was enacted. Although this act only dealt with reducing air pollution by setting emissions standards for stationary sources such as power plants and steel mills at the beginning, amendments of 1965, 1966 and 1967 focused on establishing standards for automobile emissions (AMS 2005). Emission control was first required on light-duty gasoline vehicles (LDVs) by U.S. EPA in the 1968 model year. Developed and refined over a period of more than 30 years, these controls have become more effective at reducing LDV emissions (FCAP 2004).

The relative importance of emissions from HDDVs has increased significantly because today's gasoline powered vehicles are more than 95% cleaner than vehicles in 1968. Considering that HDDVs typically have a life cycle of over one million miles, may be on the road as long as 30 years, and will continue to play a major emission inventory role with increases in goods movement with their high durability and reliability, modeling of HDDV emissions is going to become increasingly important in air quality planning.

1.2 Current Heavy-Duty Vehicle Emissions Modeling Practices

In current regional and microscale modeling conducted in every state except California, HDDV emissions rates are taken from the U.S. Environmental Protection Agency's (EPA's) MOBILE 6.2 model (U.S. EPA 2001b). MOBILE 6.2[1] emission rates were derived from baseline emission rates (gram/brakehorsepower-hour) developed in the laboratory using engine dynamometer test cycles. While different driving cycles have been developed over the years, dynamometer testing is conceptually designed to obtain a "representative sample" of vehicle operations. These work-based emission rates are then modified through a series of conversion and correction factors to obtain approximate emission rates in units of grams/mile that can be applied to on-road vehicle activity (vehicle miles traveled), as a function of temperature, humidity, altitude, average vehicle speed, etc. (Guensler 1993). The conversion process used to translate laboratory emission rates to on-road emission rates employs fuel density, brake specific fuel consumption, and fuel economy for each HDDV technology class. However, the emission rate conversion process does not appropriately account for the impacts of roadway operating conditions on brake specific fuel consumption and fuel economy (Guensler, et al. 1991).

The U.S. Environmental Protection Agency (U.S. EPA) is currently developing a new set of modeling tools for the estimation of emissions produced by on-road and off-road mobile sources. The new **Mo**tor **V**ehicle **E**missions **S**imulator, known as MOVES[2] (Koupal, et al. 2004), is a modeling system designed to better predict emissions from on-road operations. The philosophy behind MOVES is to develop a model that is as directly data-driven as possible, meaning that emission rates are developed from second-by-second or binned emission rate data.

1.3 Research Approaches and Objectives

The major effort of this research is to develop a new heavy-duty vehicle load-based modal emission rate model that overcomes some of the limitations of existing models and emission

[1]MOBILE = Current mobile source emissions model used for State Implementation Plan emission inventories.

[2]MOVES = Mobile Vehicle Emissions Estimator, next generation mobile source emissions model. The model will be used for State Implementation Plan emission inventories and will replace the current MOBILE model.

rates prediction methods. This model is part of the proposed Heavy-Duty Diesel Vehicle Modal Emission Modeling (HDDV-MEM) which was developed by Georgia Institute of Technology (Guensler, et al. 2006). HDDV-MEM differs from other proposed HDDV modal models (Barth, et al. 2004, Frey, et al. 2002, Nam 2003) in that the modeling framework first predicts second-by-second engine power demand as a function of vehicle operating conditions and then applies brake-specific emission rates to these activity predictions. This means that HDDV emission rates are predicted as a function of engine horsepower loads for different driving modes. Hence, the basic algorithm and matrix calculation in the HDDV-MEM should be transferable to MOVES. The new model implementation is similar in general structure to previous model emission rate model known as Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE[1]) model developed by Georgia Institute of Technology several years ago (Bachman 1998, Guensler, et al. 1998, Bachman, et al. 2000).

The major effort of this research consists of a number of specific objectives outlined below:

- Develop a new load-based modal emission rate model to improve spatial/temporal emissions modeling;

- Develop a HDDV modal emission rate model to more accurately estimate on-road HDDV emissions;

- Develop a modal model that can be verified at multiple levels;

- Develop a HDDV modal emission rate model that can be integrated into the MOVES.

1.4 Summary of Research Contributions

There are four major contributions developed by this research. First, a framework for emission rate modeling suitable for predicting emissions at different scales (microscale, mesoscale, and macroscale) is established. Since this model is developed using on-board emissions data which are collected under real-world conditions, this model will provide capabilities for integrating necessary vehicle activity data and emission rate algorithms to support second-by-second and link-based emissions prediction. Combined with GIS framework, this model will improve spatial/temporal emissions modeling.

[1]MEASURE = Mobile Emissions Assessment System for Urban and Regional Evaluation Model. This model is a prototype GIS-based modal emissions model.

Second, the relationship between engine power and emissions is explored and integrated into the modeling framework. Research results indicate that engine power is more powerful than surrogate variables to present load data in the proposed model. Based on the important role of engine power in explaining the variability of emissions, it is better to include the load data measurement during emission data collection procedure. Meanwhile, development of methods to simulate real world engine power is equally important.

Third, this research verifies that vehicle emission rates are highly correlated with modal vehicle activity. To get better understanding of driving modes, it is important to examine not only emission distributions, but also engine power distributions.

Finally, a dynamic framework is created for further improvement. As more databases become available, this approach could be re-run to obtain a more reliable load-based modal emission model based on the same philosophy.

1.5 Report Organization

Chapter 2 examines the diesel fuel combustion process and its relationship to diesel engine emissions formation. Chapter 3 overviews the existing heavy-duty vehicle emission models and presents the proposed heavy-duty diesel vehicle modal emission model (HDDV-MEM). Chapter 4 provides an overview of the emission rate testing databases provided by U.S. EPA, the quality assurance and quality control (QA/QC) procedures to review the validity of the data, and the methods used to post-process these databases to correct data deficiencies. In Chapter 5, the various statistical models considered for data analysis are discussed. Chapter 6 selects the database used to develop the conceptual model and discusses the influence of explanatory variables on emissions. Chapter 7 covers sensitivity tests of driving mode definitions and outlines the potential impacts on derived models. Chapters 8 to 11 elaborate the different emission models developed for idle, deceleration, acceleration and cruise driving modes. In Chapter 12, research results are verified. Finally, Chapter 13 presents a discussion and conclusion on research results.

CHAPTER 2

2. HEAVY-DUTY DIESEL VEHICLE EMISSIONS

Diesel engines differ from gasoline engines in terms of the combustion processes and engine size, giving rise to their different emission properties and therefore different emissions standards. This chapter examines the diesel fuel combustion process and its relationship to diesel engine emissions formation followed by a summary of the emission regulations for diesel engines.

2.1 How Diesel Engine Works

By far the predominant engine design for transportation vehicles is the reciprocating internal combustion (IC) engine which operates either on a four-stroke or a two-stroke cycle. The two-stroke engine is commonly found in lower-power applications such as snowmobiles, lawnmowers, mopeds, outboard motors and motorcycles, while both gasoline and diesel automotive engines are classified as four-stroke engines. To understand the formation and control of emissions, it is necessary to first develop an understanding of the operation of the internal combustion engine.

2.1.1 The Internal Combustion Engine

Internal combustion engines generate power by converting the chemical energy stored in fuels into mechanical energy. The engine is termed "internal combustion" because combustion occurs in a confined space called a combustion chamber. Combustion of the fuel charge inside a chamber causes a rapid rise in temperature and pressure of the gases in the chamber, which are permitted to expand. The expanding gases are used to move a piston, turbine blades, rotor, or the engine itself.

The four-stroke gasoline engine cycle is also called Otto cycle, in honor of Nikolaus Otto, who is credited with inventing the process in 1867. The four piston strokes are illustrated in Figure 2-1. The following processes take place during one cycle of operation:

1. **Intake stroke:** the piston starts at the top, the intake valve opens, and the piston moves down to let the engine take in a fresh charge composed of a mixture of fuel and air (for spark-ignition or gasoline engine) or air only (for auto-ignition or diesel engine). (Part 1 of the figure.)

2. **Compression stroke:** then the piston moves back up to compress this fuel/air mixture (gasoline engines) or the air only (diesel engines). In gasoline engines combustion is started by ignition from a spark plug, in diesel engines auto-ignition occurs when fuel is injected into the compressed air which has achieved a high temperature through compression such that the temperature is high enough to cause self-ignition. (Part 2 of the figure.)

3. **Expansion stroke:** when the piston reaches the top of its stroke, the combustion process results in a substantial increase in the gas temperature and pressure and drives the piston down. (Part 3 of the figure.)

4. **Exhaust stroke:** once the piston hits the bottom of its stroke, the exhaust valve opens and the exhaust leaves the cylinder into the exhaust manifold and then into the tail pipe. Discharge of the burnt gases (exhaust) from the cylinder occurs to make room for the next cycle. (Part 4 of the figure.)

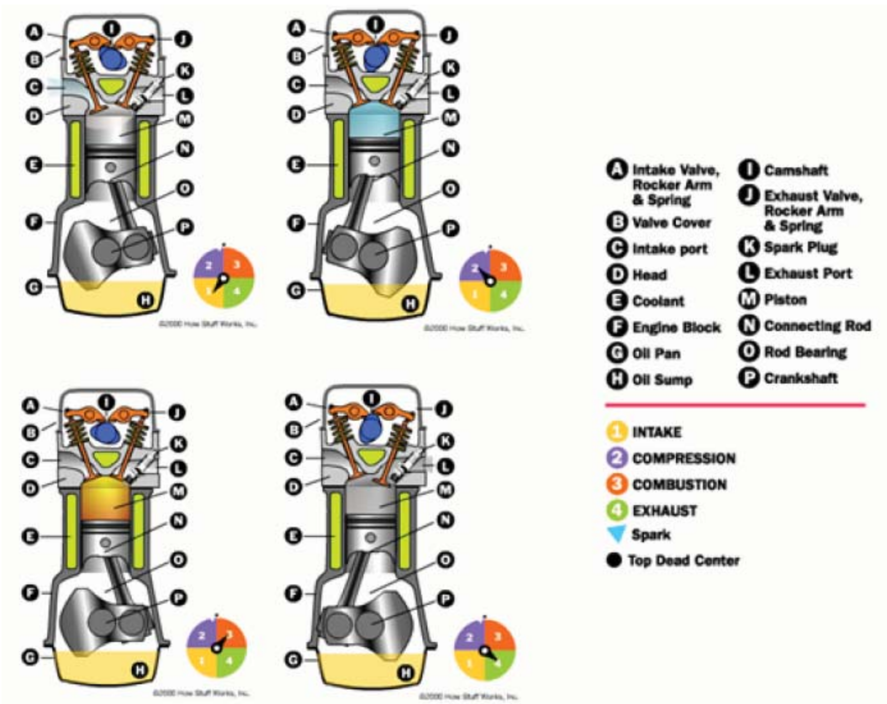


Figure 2.1 Actions of a four-stroke gasoline internal combustion engine -- Adapted from (HowStuff-Works 2005)

Figure 2-1 is a diagrammatic representation of the four strokes of an internal combustion engine. The upper end of the cylinder consists of a clearance space in which ignition and combustion occur. The expanding medium pushes against the piston head inside the cylinder, causing the piston to move; this straight line motion of the piston is converted into the desired rotary motion of the wheels by means of a drivetrain consisting of a connecting rod and crankshaft. Figure 2-1 illustrates that the only stroke that delivers useful work is the expansion stroke; the other three strokes are thus termed idle strokes. The reader interested in a detailed description of the internal combustion engine is referred to specialized texts, such as Heywood (Heywood 1998) and Newton et al. (Newton, et al. 1996).

2.1.2 Comparison with the Gasoline Engine

The diesel engine employs the compression ignition cycle. German engineer Rudolf Diesel developed the idea for the diesel engine and received the patent on February 23, 1893. His goal was to create an engine with high efficiency. Figure 2-2 is a diagrammatic representation of the four strokes of a diesel engine. The main differences between the gasoline engine and the diesel engine are:

- A gasoline engine compresses at a ratio of 8:1 to 12:1, while a diesel engine compresses at a ratio of 14:1 to as high as 25:1. The higher compression ratio of the diesel engine leads to higher peak combustion temperatures and better fuel efficiency.

- Unlike a gasoline engine, which takes in a mixture of gas and air, compresses it and ignites the mixture with a spark, a diesel engine takes in just air, compresses it and then injects fuel into the compressed air. The heat of the compressed air spontaneously ignites the fuel.

- Gasoline engines generally use either carburetion, in which the air and fuel is mixed long before the air enters the cylinder, or port fuel injection, in which the fuel is injected just prior to the intake stroke (outside the cylinder), while diesel engines use direct fuel injection – the diesel fuel is injected directly into the cylinder.

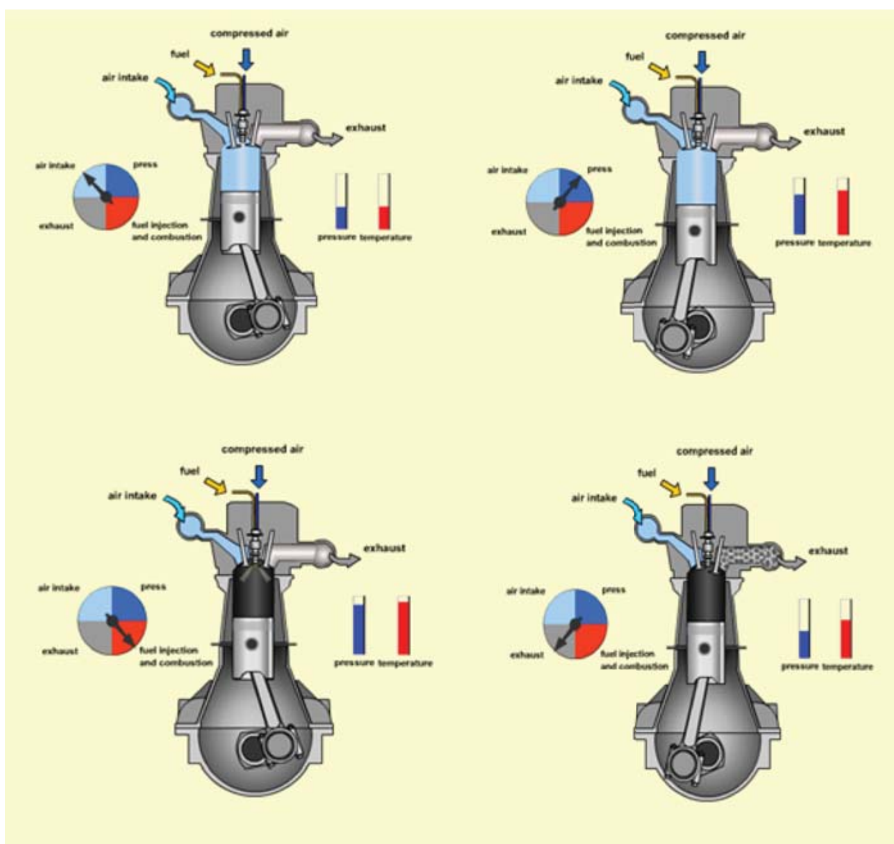


Figure 2.2 Actions of a four-stroke diesel engine (HowStuffWorks 2005)

2.2 Diesel Engine Emissions

Like any other internal combustion engine, diesel engines convert the chemical energy contained in diesel fuel into mechanical power. Diesel fuel is injected under pressure into the engine cylinder, where it mixes with air and combustion occurs. Diesel fuel is heavier and oilier than gasoline. Diesel fuel evaporates much more slowly than gasoline, with a boiling point that is actually higher than that of water. The lean nature of the diesel-air mixture results in a combustion environment that produces lower emission rates of carbon monoxide (CO) and hydrocarbons (HC) compared to gasoline-powered engines. However, diesel engines do produce relatively high level emissions of oxides of nitrogen (NO_x) and particulate matter (PM), especially fine particulate matter. This section will discuss oxides of nitrogen and particulate emissions in detail.

2.2.1 Oxides of Nitrogen and Ozone Formation

Oxides of nitrogen, a mixture of nitric oxide (NO) and nitrogen dioxide (NO_2), are produced from the destruction of atmospheric nitrogen (N_2) during the combustion process. Atmospheric air generally consists of 80% N_2 and 20% O_2, and these elements are stable because of the moderate temperatures and pressures. However, during high temperature and pressure conditions of combustion, excess oxygen in the combustion chamber reacts with N_2 to create NO which is quickly transformed into NO_2. The role of nitrogen contained in the air in NO formation was initially postulated by Zeldovich (Zeldovich, et al. 1947). In near-stoichiometric or lean systems the mechanisms associated with NO formation (as many as 30 or so independent chemical reactions that also involve participation of hydrocarbon species) can generally be simplified to the following:

Reaction 1: $O_2 \rightleftharpoons O + O$

Reaction 2: $O + N_2 \rightleftharpoons NO + N$

Reaction 3: $N + O_2 \rightleftharpoons NO + O$

In near-stoichiometric and fuel-rich mixtures, where the concentration of OH radicals can be high, the following reaction also takes place:

Reaction 4: $N + OH \rightleftharpoons NO + H$

Reaction 4, together with reactions 1, 2 and 3, are known as the extended Zeldovich mechanism. It is also important to note that emitted nitric oxide (NO) will oxidize to nitrogen dioxide (NO_2) in the atmosphere over a period of a few hours.

Oxides of nitrogen (NO_X) are reactive gases that cause a host of environmental concerns impacting adversely on human health and welfare. Nitrogen dioxide (NO_2), in particular, is a brownish gas that has been linked with higher susceptibility to respiratory infection, increased airway resistance in asthmatics, and decreased pulmonary function. Most importantly, NO_X emitted from heavy-duty vehicles plays a major role in the formation of ground level ozone pollution, which causes wide-ranging damage to human health and the environment (U.S. EPA 1995). Ozone is a colorless, highly reactive gas with a distinctive odor. Naturally, ozone is formed by electrical discharge (lightning) and in the upper atmosphere at altitudes between 15 and 35 km. Stratospheric ozone protects the Earth from harmful ultraviolet radiation from the sun. However, ground level ozone is formed by chemical reactions involving NO_X and volatile organic compounds (VOCs) combining in the presence of heat and sunlight. These two categories of pollutants are also referred to as ozone precursors. The production of photochemical oxidants usually occurs over several hours which means that the highest concentrations of ozone normally occur on summer afternoons, in areas downwind of major sources of ozone precursors. The simplified reaction processes are illustrated as:

$$NO_2 + VOC + \text{sunlight (UV)} \Rightarrow NO_2 + O_2 + \text{sunlight (UV)} \Rightarrow NO + O_3$$

At ground level, elevated ozone concentrations can cause health and environmental problems. Ozone can affect the human cardiac and respiratory systems, irritating the eyes, nose, throat, and lungs. Symptoms of ozone exposure include itchy and watery eyes, sore throats, swelling within the nasal passages and nasal congestion. Effects from ozone are experienced only for the period of exposure to elevated levels. EPA promulgated 8-hour ozone standards in 1997 and designated an area as nonattainment if it has violated, or has contributed to violations of, the national 8-hour ozone standard over a three-year period.

2.2.2 Fine Particulate Matter ($PM_{2.5}$)

Particulate matter (PM) is a complex mixture of solid and liquid particles (excluding water) that are suspended in air. These particles typically consist of a mixture of inorganic and organic chemicals, including carbon, sulfates, nitrates, metals, acids, and semivolatile compounds. The size of PM in air ranges from approximately 0.005 to 100 micrometers (μm) in aerodynamic diameter -- the size of just a few atoms to about the thickness of a human hair. U.S. EPA defined three general categories for PM as coarse (10 to 2.5 μm), fine (2.5 μm or smaller), and ultrafine (0.1 μm or smaller).

Heavy-duty diesel vehicles are known to emit large quantities of small particles (Kittelson, et al. 1978). A majority of the PM found in diesel exhaust is in the nanometer size range.

Lloyd found that more than 90% of fine particles from heavy-duty vehicles are smaller than 1μm in diameter (Lloyd and Cackette 2001).

Fine PM can cause not only human health problems and property damage, but also adversely impact the environment through visibility reduction and retard plant growth (Davis, et al. 1998). Health studies have shown a significant association between exposure to fine particles and premature death from heart or lung diseases. Other important effects include aggravation of respiratory and cardiovascular disease, lung disease, decreased lung function, or asthma attacks. Individuals particularly sensitive to fine particle exposure include older adults, people with heart and lung disease, and children (U.S. EPA 2005). EPA promulgated the PM2.5 standard in 1997 and included a 24-hour standard for PM2.5 set at 65 micrograms per cubic meter (μg/m3), and an annual standard of 15 μg/m3.

2.3 Heavy-Duty Diesel Vehicle Emission Regulations

2.3.1 National Ambient Air Quality Standards

The Clean Air Act, which was last amended in 1990, requires the U.S. EPA to set National Ambient Air Quality Standards (NAAQS) to safeguard public health against six common air pollutants: ozone (O_3), particulate matter (PM), sulfur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2) and lead (Pb). The Clean Air Act established two types of national air quality standards. Primary standards set limits to protect public health, including the health of "sensitive" populations such as asthmatics, children, and the elderly. Secondary standards set limits to protect public welfare, including protection against decreased visibility, damage to animals, crops, vegetation, and buildings (CFR 2004a). Table 2-1 illustrates the current NAAQS for ambient concentrations of various pollutants. Units of measure for the standards are parts per million by volume (ppmv), milligrams per cubic meter of air (mg/m^3), and micrograms per cubic meter of air ($\mu g/m^3$).

Table 2-1. National Ambient Air Quality Standards (U.S. EPA 2006)

Pollutant	Average Times	Standard Value	Standard Type
Carbon Monoxide (CO)	8-hour Average	9 ppmv (10 mg/m^3)	Primary
	1-hour Average	35 ppmv (40 mg/m^3)	Primary
Nitrogen Dioxide (NO_2)	Annual Arithmetic Mean	0.053 ppmv (100 $\mu g/m^3$)	Primary & Secondary
Ozone (O_3)	1-hour Average	0.12 ppmv (235 $\mu g/m^3$)	Primary & Secondary
	8-hour Average	0.08 ppmv (157 $\mu g/m^3$)	Primary & Secondary

Pollutant	Average Times	Standard Value	Standard Type
Lead (Pb)	Quarterly Average	1.5 μg/m^3	Primary & Secondary
Particulate (PM10)	Annual Arithmetic Mean	50 μg/m^3	Primary & Secondary
	24-hour Average	150 μg/m^3	Primary & Secondary
Particulate (PM2.5)	Annual Arithmetic Mean	15 μg/m^3	Primary & Secondary
	24-hour Average	65 μg/m^3	Primary & Secondary
Sulfur Dioxide (SO_2)	Annual Arithmetic Mean	0.030 ppmv (80 μg/m^3)	Primary
	24-hour Average	0.14 ppmv (365 μg/m^3)	Primary
	3-hour Average	0.50 ppmv (1300 μg/m^3)	Secondary

2.3.2 Heavy-Duty Engine Certification Standards

Heavy-duty vehicles are defined as vehicles of GVWR (gross vehicle weight rating) above 8,500 lbs in the federal jurisdiction and above 14,000 lbs in California (model year 1995 and later). Diesel engines used in heavy-duty vehicles are further divided into service classes by GVWR, as follows:

- Light heavy-duty diesel engines: 8,500<LHDDE<19,500 (14,000<LHDDE<19,500 in California, 1995+)

- Medium heavy-duty diesel engines: 19,500≤MHDDE≤33,000

- Heavy heavy-duty diesel engines (including urban bus): HHDDE>33,000

Under the federal light-duty Tier 2 regulation (phased in beginning 2004), vehicles of GVWR up to 10,000 lbs used for personal transportation have been re-classified as "medium-duty passenger vehicles" (MDPV – primarily larger SUVs and passenger vans) and are subject to the light-duty vehicle legislation. Thus, the same diesel engine model used for the 8,500-10,000 lbs vehicle category may be classified as either light- or heavy-duty and certified to different standards, depending on the manufacturer-defined application (CFR 2004b). Except for the heavy-duty vehicles classified as LDVs, all heavy-duty vehicle emissions standards are established using the engine dynamometer certification process.

2.3.3 Heavy-Duty Engine Emission Regulations

EPA regulates heavy-duty vehicle emissions for compliance with emissions standards over the useful life of the engine. Useful life is defined as follows (U.S. EPA and California) (CFR 2004c):

LHDDE – 8 years/110,000 miles (whichever occurs first)
MHDDE – 8 years/185,000 miles
HHDDE – 8 years/290,000 miles

Federal useful life requirements were later increased to 10 years, with no change to the above mileage numbers, for the urban bus PM standard (1994+) and for the NOx standard (1998+). The emission warranty period is 5 years/100,000 miles (5 years/100,000 miles/3,000 hours in California), but no less than the basic mechanical warranty for the engine family. Table 2-2 shows the heavy-duty engine emissions standards by model year group.

Table 2-2. Heavy-Duty Engine Emissions Standards (U.S. EPA 1997)

Year	HC (g/bhp-hr)	CO (g/bhp-hr)	NO_x (g/bhp-hr)	PM (g/bhp-hr)
Heavy-Duty Diesel Truck Engines				
1988	1.3	15.5	10.7	0.60
1990	1.3	15.5	6.0	0.60
1991	1.3	15.5	5.0	0.25
1994	1.3	15.5	5.0	0.10
1998	1.3	15.5	4.0	0.10
Urban Bus Engines				
1991	1.3	15.5	5.0	0.25
1993	1.3	15.5	5.0	0.10
1994	1.3	15.5	5.0	0.07
1996	1.3	15.5	5.0	0.05*
1998	1.3	15.5	4.0	0.05*
* in-use PM standard 0.07				

2.4 Heavy-Duty Diesel Vehicle Emission Modeling

There are several models currently used to estimate emissions from heavy-duty vehicles. The most common emission rate models are VMT-based or cycle-based models, developed from laboratory test facility driving cycle data. Due to lack of available data representing real world conditions, all previous models were developed based upon engine dynamometer data. The following chapter will address this issue in detail.

CHAPTER 3

3. HEAVY-DUTY DIESEL VEHICLE EMISSIONS MODELING

Several models are currently used to estimate emissions from heavy-duty vehicles. A comprehensive review of the existing heavy-duty vehicle emission models will help modelers understand the different approaches and how they can contribute to the development of enhanced emission rate modeling techniques.

The most common emission rate models are VMT-based or cycle-based developed from laboratory test facility driving cycle data. Fuel-based models model emissions as a function of fuel usage rate as well as other parameters. In the 1990s, even the proposed enhanced modal models, designed to predict emissions as a function of speed and acceleration profiles of vehicles, were still based upon statistical analysis of cycle-based data (Bachman 2000; Fomunung 2000). More recent emission rate modeling frameworks are proposing to model modal emission rates on a second-by-second basis directly from the vehicle operating mode.

3.1 VMT-Based Vehicle Emission Models

The current emission rate models used by state and federal agencies include the Mobile Source Emission Model (MOBILE) series of models developed by the U.S. Environmental Protection Agency (U.S. EPA) and the Emission Factor Emission Inventory Model (EMFAC) series developed by California Air Resources Board (CARB).

3.1.1 MOBILE

MOBILE (U.S. EPA 1993), developed by the US EPA in the late 1970s to estimate vehicle emission, has since become the nation's standard in assessing the emission impacts of various transportation inputs. MOBILE uses the method of base emission rates and correction factors. This model has undergone significant expansion and improvements over the years. The latest version is MOBILE6 released in February 2002 (U.S. EPA 2002a).

MOBILE is based on engine dynamometer test data from selected driving cycles. The Federal Test Procedure (FTP) transient cycle is composed of a unique profile of stops, starts, constant speed cruises, accelerations and decelerations. Different driving cycles are developed to simulate both urban and freeway driving. A concern with driving cycles is that they may not be sufficiently representative of real-world emissions (Kelly and Groblicki 1993; Denis et al. 1994). For HDV emission rates, MOBILE uses the method of base emission rates and conversion factors which convert the g/bhp-hr emissions estimates observed in the laboratory to g/mile emission rates, to be consistent with available travel information. Conversion factors are used to convert the g/bhp-hr emissions estimates to grams per mile traveled. These conversion factors contribute a large source of uncertainty to the MOBILE model since the BSFC (brake specific fuel consumption) data are aggregated for the fleet and may not represent in-use vehicle characteristics (Guensler et al. 1991). Conversion factors have improved accuracy in MOBILE6 due to improved data, but fundamental flaws remain (Guensler et al. 2006).

3.1.1.1 Diesel Engine Test Cycles

EPA currently uses the transient Federal Test Procedure (FTP) engine dynamometer cycle, which includes both engine cold and warm start operations, for heavy-duty vehicles (CFR Title 40, Part 86.1333). Unlike the chassis dynamometer test for light-duty vehicle, the engine is removed from the vehicle's chassis, mounted on the engine dynamometer test stands, and operated in the transient FTP test cycle. The transient cycle (Figure 3-1) consists of four phases: the first is a NYNF (New York Non Freeway) phase typical of light urban traffic with frequent stops and starts, the second is LANF (Los Angeles Non Freeway) phase typical of crowded urban traffic with few stops, the third is a LAFY (Los Angeles Freeway) phase simulating crowded expressway traffic in Los Angeles, and the fourth phase repeats the first NYNF phase. This cycle consists of a cold start after parking overnight, followed by idling, acceleration and deceleration phases, and a wide variety of different speeds and loads sequenced to simulate the running of the vehicle that corresponds to the engine being tested. There are few stabilized running conditions, and the average load factor is about 20 to 25% of the maximum horsepower available at a given speed.

Emission and operation parameters are measured while the engine operates during the test cycle. The engine torque is determined by applying performance percentages with an engine lug curve (maximum torque curve). Engine torque is then converted to engine brake horsepower using engine revolution per minute (RPM). Brake specific emissions rates are reported in g/bhp-hr and then converted to g/mile using pre-defined conversion factors (CFR Title 40, Part 86.1342-90).

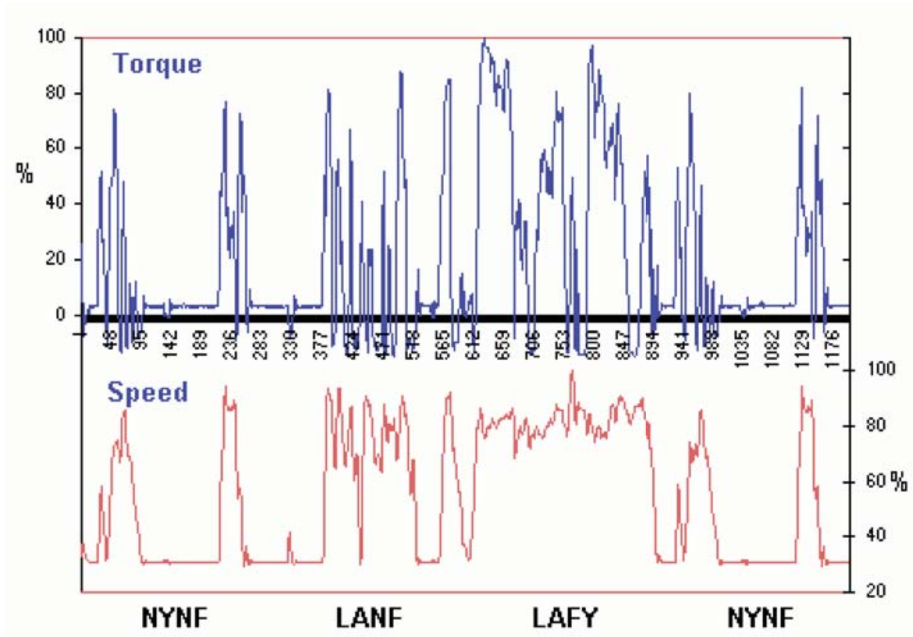


Figure 3-1 FTP Transient Cycle (DieselNet 2006)

Because the engine dynamometer test procedure does not directly account for the impacts from load and grade changes, a chassis dynamometer test procedure and the cycle known as the HDV urban dynamometer driving schedule (HDV-UDDS) was developed [CFR Title 40, Part 86, App. I], sometimes referred to as "cycle D". This cycle is different from the UDDS cycle for light-duty vehicles (FTP-72). This HDV cycle lasts 1060 seconds and covers 5.55 miles. The average speed for HDV UDDS is 18.86 mph while the maximum speed is 58 mph. Figure 3-2 shows the speed profile for the chassis UDDS test.

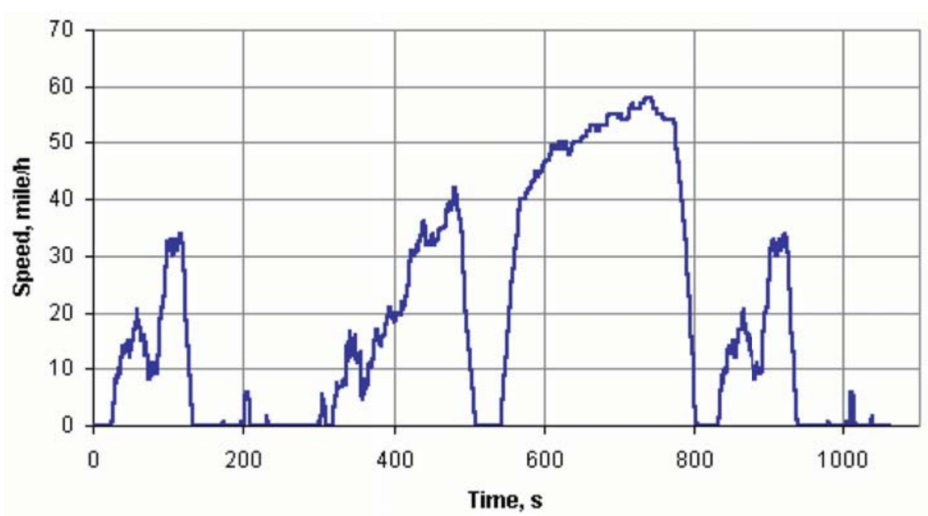


Figure 3-2 Urban Dynamometer Driving Schedule Cycle for Heavy-Duty Vehicle (DieselNet 2006)

3.1.1.2 Baseline Emission Rates

Baseline emission rates (g/bhp-hr) for heavy-duty vehicles are obtained from the engine dynamometer test results collected during U.S. EPA's cooperative test program with engine manufacturers. The zero mile levels and deterioration rates for NO_x, CO, and HC are presented in the following tables for heavy-duty gasoline and diesel engines. All the emission rates are available from "Update of Heavy-Duty Emission Levels (Model Years 1998-2004+) for Use in MOBILE6" (Lindhjem and Jackson 1999).

Table 3-1. Heavy-Duty Vehicle NO_x Emission Rates in MOBILE6

Model Year Class	Zero Mile Level (g/bhp-hr)				Deterioration (g/bhp-hr/10,000 miles)			
	Gasoline Engine	Diesel Engine			Gasoline Engine	Diesel Engine		
		Heavy	Med.	Light		Heavy	Med.	Light
1988-1989	4.96	6.28	6.43	4.34	0.044	0.01	0.009	0.002
1990	3.61	4.85	4.85	4.85	0.026	0.004	0.006	0.011
1991-1993	3.24	4.56	4.53	1.38	0.038	0.004	0.007	0.003
1994-1997	3.24	4.61	4.61	1.08	0.038	0.003	0.001	0.001
1998-2003	2.59	3.68	3.69	3.26	0.038	0.003	0.001	0.001
2004+	2.59	1.84	1.84	1.63	0.038	0.003	0.001	0.001

Table 3-2 Heavy-Duty Vehicle CO Emission Rates in MOBILE6

Model Year Class	Zero Mile Level (g/bhp-hr)				Deterioration (g/bhp-hr/10,000 miles)			
	Gasoline Engine	Diesel Engine			Gasoline Engine	Diesel Engine		
		Heavy	Med.	Light		Heavy	Med.	Light
1988-1989	13.84	1.34	1.70	1.21	0.246	0.008	0.018	0.022
1990	6.89	1.81	1.81	1.81	0.213	0.005	0.007	0.012
1991-1993	7.10	1.82	1.26	0.40	0.255	0.003	0.010	0.004
1994-1997	7.10	1.07	0.85	1.19	0.255	0.004	0.009	0.003
1998-2003	7.10	1.07	0.85	1.19	0.255	0.004	0.009	0.003
2004+	7.10	1.07	0.85	1.19	0.255	0.004	0.009	0.003

Table 3-3 Heavy-Duty Vehicle HC Emission Rates in MOBILE6

Model Year Class	Zero Mile Level (g/bhp-hr)				Deterioration (g/bhp-hr/10,000 miles)			
	Gasoline Engine	Diesel Engine			Gasoline Engine	Diesel Engine		
		Heavy	Med.	Light		Heavy	Med.	Light
1988-1989	0.62	0.47	0.66	0.64	0.023	0.001	0.002	0.002
1990	0.35	0.52	0.52	0.52	0.023	0.000	0.001	0.001
1991-1993	0.33	0.30	0.40	0.47	0.021	0.000	0.001	0.001
1994-1997	0.33	0.22	0.31	0.26	0.021	0.001	0.001	0.001
1998-2003	0.33	0.22	0.31	0.26	0.021	0.001	0.001	0.001
2004+	0.33	0.22	0.31	0.26	0.021	0.001	0.001	0.001

3.1.1.3 Conversion Factors

Because emission standards for both gasoline and diesel heavy-duty vehicles are expressed in terms of grams per brake-horsepower hour (g/bhp-hr), the MOBILE6.2 model employs conversion factors of brake horsepower-hour per mile (bhp-hr/mile) to convert the emission certification data from engine testing to grams per mile. Conversion factors are a function of fuel density, brake-specific fuel consumption (BSFC), and fuel economy for each HDV class (U.S. EPA 2002b). The conversion factors were calculated using Equation 3-1:

$$\text{Conversion Factor (bhp-hr/mi)} = \frac{\text{Fuel Density (lb/gal)}}{\text{BSFC (lb/bhp-hr)} \times \text{Fuel Economy (mi/gal)}} \qquad \text{(Equation 3-1)}$$

To calculate BSFC, U.S. EPA first obtained data from model year 1987 through 1996 supplied by six engine manufacturers (U.S. EPA 2002d). U.S. EPA then performed regression analysis for BSFCs by model year for each weight class and used a logarithmic curve to extrapolate values prior to 1988 and after 1995, since sales data were only available for model years 1988 through 1995 (U.S. EPA 2002d).

Fuel economy was calculated using a regression curve derived from the 1992 Truck Inventory and Use Survey (TIUS) conducted by the U.S. Census Bureau. Fuel densities were determined from National Institute for Petroleum and Energy Research (NIPER) publications for both gasoline and diesel (Browning 1998). Using the equation defining the conversion factor together with the data described above, weight class specific conversion factors were calculated for gasoline and diesel vehicles for model years 1987 through 1996 (U.S. EPA 2002c).

3.1.2 EMFAC

EMFAC (CARB 2007) was developed by CARB separately from MOBILE based upon the presence of vehicle technologies in the on-road fleet that would be subject to more stringent standards and fuels used in California. The latest version, EMFAC 2002, was released in September 2002. EMFAC can estimate emissions for calendar years 1970 to 2040.

EMFAC abandoned the use of conversion factors from EMFAC 2000 and used chassis dynamometer data collected for 70 trucks tested over the Urban Dynamometer Driving Schedule (UDDS). Although the use of UDDS test data marked a significant improvement, it is hard to say that UDDS adequately represented the full range of heavy duty diesel operation. Although the cycle was constructed from actual truck activity data, it lacks extended cruises known to cause many trucks to default to a high NO_x emitting, fuel saving mode referred to as “Off-Cycle”

NOx. The cycle also lacks hard accelerations known to result in high emissions of particulate matter (CARB 2002).

CARB continues to develop more mode test cycles designed to better depict the emissions of HDDVs under real world conditions, including emissions from engine programming to go “off-cycle” at certain speeds. Activity data from instrumented truck studies conducted by Battelle and Jack Faucett Associates for CARB (CARB 2002) have been used to develop a four mode heavy-heavy-duty diesel cycle. Figure 3-3 shows these four mode cycles developed by CARB. The creep mode produced the greatest gram per mile results followed by the transient and the cruise mode. The transient and cruise modes produced higher and lower emissions, respectively, than the UDDS (CARB 2002).

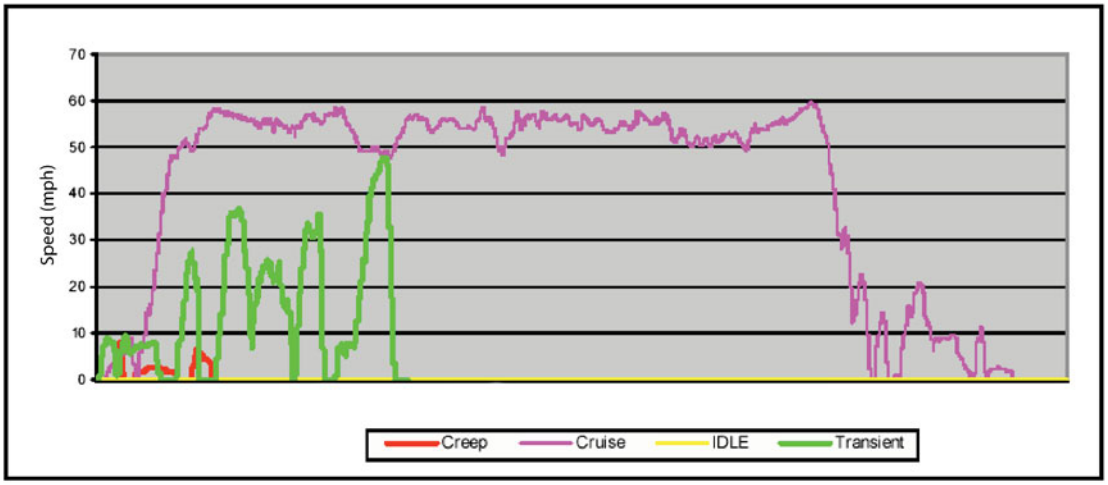


Figure 3-3 CARB’s Four Mode Cycles (CARB 2002)

3.1.3 Summary

EPA’s MOBILE series models have significantly improved through the series of model revisions from 1970s. However, the MOBILE series of models still has major modeling defects for the heavy-duty components. These defects have been widely recognized for more than 10 years (Guensler et al. 1991). One of the most frequently stated defects is that fleet average speed, which aggregates other vehicle activity factors that may yield significant bias in emissions characterization, is used to characterize vehicle emission rates.

In developing emissions inventories using the MOBILE and EMFAC (CARB 2007) emission rate models, vehicle activity is estimated using travel demand models. The estimation of VMT was based on EPA’s fleet characterization study (U.S. EPA 1998). It is common to estimate heavy-duty travel as a fixed percentage of predicted traffic volumes (TRB 1995). This estimate is not correct since heavy-duty truck travel does not follow the same spatial and temporal patterns as light-duty vehicle travel (Schlappi et al. 1993).

3.2 Fuel-Based Vehicle Emission Models

The fuel-based emission inventory models for heavy-duty diesel trucks combine vehicle activity data (i.e., volume of diesel fuel consumed) with emission rates normalized to fuel consumption (i.e., mass of pollutant emitted per unit volume of fuel burned) to estimate emissions within a region of interest (Dreher and R. Harley 1998). This approach was proposed to increase accuracy of truck VMT estimation by combining state level truck VMT with statewide fuel sales to estimate total heavy-duty truck activity, using the amount of fuel consumed as a measure of activity.

In California, fuel consumption data are available through tax records at the statewide level and this statewide fuel consumption can be apportioned to provide emission estimates for an individual air basin by month, day of week, and time of day. At the same time, emission rates are normalized to fuel consumption using Equation 3-2:

$$EI_p = \frac{S_p}{BSFC} \qquad \text{(Equation 3-2)}$$

where EI_p: emission index for pollutant P, in units of mass of pollutant emitted per unit mass of fuel burned;

S_p : brake specific pollutant emission rate obtained from the dynamometer test, expressed in g/bhp-hr units;

BSFC : brake specific fuel consumption of the engine being tested, also in g/bhp.

Exhaust emissions are estimated by multiplying vehicle activity, as measured by the volume of fuel used, by emission rates which are normalized to fuel consumption and expressed as grams of pollutant emitted per gallon of diesel fuel burned instead of grams of pollutant per mile (Dreher and R. Harley 1998). Average emission rates for subgroups of vehicles are weighted by the fraction of total fuel used by each vehicle subgroup to obtain an overall fleet-average emission rate. The fleet-average emission rate is multiplied by regional fuel sales to compute pollutant emissions (Singer and Harley 1996).

The advantages of the fuel-based approach include the fact that fuel-use data are available from tax records in California. Furthermore, emission rates normalized to fuel consumption vary considerably less over the full range of driving conditions than travel-normalized emission factors (Singer and Harley 1996). The disadvantage is obvious, too. Tax records are not available for other states. It is difficult to get input data outside of California, limiting the scope of the modeling approach. Furthermore, the users first have to run two models to predict fuel used and then predict emission rates, which is not statistically efficient.

3.3 Modal Emission Rate Models

Modal emission rate models work on the premise that emissions are better modeled as a function of specific modes of vehicle operation (idle, steady-state cruise, various levels of acceleration/deceleration, etc.), than as a function of average vehicle speed (Bachman 1998; Ramamurthy et al. 1998; U.S. EPA 2001b). Emissions of heavy-duty vehicles powered by diesel cycle engines are more likely to be a function of brake work output of engine than normal gasoline vehicles, because instantaneous emissions levels of diesel engine are highly correlated with the instantaneous work output of the engine (U.S. EPA 2001b).

With the consideration of vehicle modal activity, EPA and various research communities have been developing modal activity-based emission models. The report published by National Research Council (NRC 2000) comprehensively reviewed the modeling of mobile source emissions and provided recommendations for the improvement of future mobile source emission models. The following sections will introduce the most representative modal emission models one by one.

3.3.1 CMEM

The Comprehensive Modal Emissions Model (CMEM) (Barth et al. 2000) was developed by the Center for Environmental Research and Technology at University of California Riverside (UCR-CERT). Development of CMEM was first funded by National Cooperative Highway Research Program Project (1995-2000) and then is being enhanced and improved with EPA funding (2000-present). From 2001, CE-CERT created a modal-based inventory at the micro- (intersection), meso- (highway link), and macro- (region) scale levels for light-duty vehicles (LDV) and heavy-duty diesel (HDD) vehicles. The CMEM model derives a fuel rate from road-load and a simple powertrain model. Emissions rates are then derived empirically from the fuel rate. Fuel rate, or fuel consumption per unit time, forms the basis for CMEM.

The CMEM HDD emissions model (Barth et al. 2004) accepted the same approach as the light-duty vehicle model. In that model, second-by-second tailpipe emissions are modeled as the product of three components: fuel rate (FR), engine-out emission indices (grams of emissions/gram of fuel), and an emission after-treatment pass fraction. The model is composed of six modules: 1) engine power demand; 2) engine speed; 3) fuel-rate; 4) engine control unit; 5) engine-out emissions; and 6) after-treatment pass fraction. The vehicle power demand is determined based on operating variables [second-by-second vehicle speed (from which acceleration can be derived; note that acceleration can be input as a separate input variable), grade, and accessory use (such as air conditioning)] and specific vehicle parameters (vehicle mass, engine displacement, cross-sectional area, aerodynamics, vehicle accessory load, transmission efficiency, and drive-train efficiency, and so on). The core of the model is the fuel rate calculation which is a function of power demand and engine speed. Engine speed is determined based on vehicle velocity, gear shift schedule and power demand (Barth et al. 2004). The model uses a total of 35 parameters to estimate vehicle tailpipe emissions.

3.3.2 MEASURE

The Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE) (Bachman et al. 2000) model was developed by Georgia Institute of Technology in the late 1990s. The MEASURE model is developed within a geographic information system (GIS) and employs modal emission rates, varying emissions according to vehicle technologies and modal operation (cruise, acceleration, deceleration, idle). The model emission rate database consists of more than 13,000 laboratory tests conducted by the EPA and CARB using standardized test cycle conditions and alternative cycles (Bachman 1998). The aggregate modal model within MEASURE employs emission rates based on theoretical engine-emissions relationships. The relationships are dependent on both modal and vehicle technology variables, and they are "aggregate" in the sense that they rely on bag data to derive their modal activities (Washington et al. 1997a). Emission rates were statistically derived from the emission rate data as a function of operating mode power demand surrogates. The model uses statistical techniques to predict emission rates using a process that utilizes the best aspects of hierarchical tree-based regression (HTBR) and ordinary least squares regression (OLS) (Breiman et al. 1984). HTBR is used to reduce the number of predictor variables to a manageable number, and to identify useful interactions among the variables; then OLS regression techniques are applied until a satisfactory model is obtained (Fomunung et al. 2000). Vehicle activity variables include average speed, acceleration rates, deceleration rates, idle time, and surrogates for power demand. The MEASURE model for light-duty vehicles was completed in 2000.

MEASURE provides the following benefits since it has been developed under the GIS platform (Bachman et al. 2000): 1) manages topographical parameters that affect emissions; 2) calculates emissions from vehicle modal activities; 3) allows a 'layered' approach to individual vehicle activity estimation; and 4) aggregates emission estimates into grid cells for use in photochemical air quality models.

3.3.3 MOVES

To keep pace with new analysis needs, modeling approaches, and data, the U.S. EPA's Office of Transportation and Air Quality (OTAQ) is developing a modeling system termed MOVES (Koupal et al. 2004, U.S. EPA 2001a). This new system will estimate emissions for on-road and non-road sources, cover a broad range of pollutants, and allow multiple scale analysis, from fine-scale analysis to national inventory estimation. In the future, MOVES will serve as the replacement for MOBILE6 and NONROAD (U.S. EPA 2001a). This project was previously known as the New Generation Mobile Source Emissions Model (NGM) (U.S. EPA 2001a).

The current plan for MOVES is to use vehicle specific power (VSP) as a variable on which emission rates can be based (Koupal et al. 2002). The VSP approach to emissions characterization was developed by Jimenez-Palacios (Jimenez-Palacios 1999). VSP is a function of speed, acceleration, road grade, etc., as shown in Equation 3-3:

$$VSP = v \times (a \times (1+\varepsilon) + g \times grade + g \times C_R) + 0.5\rho \times C_D \times A \times v^3 / m \quad \text{(Equation 3-3)}$$

where:

	v:	vehicle speed (assuming no headwind) (m/s)
	a:	vehicle acceleration (m/s^2)
	ε:	mass factor accounting for the rotational masses (~0.1) - constant
	g:	acceleration due to gravity (m/s^2)
	grade:	road grade (ratio of rise to run)
	C_R:	rolling resistance (~0.0135)
	μ:	air density (1.2)
	C_D:	aerodynamic drag coefficient (dimensionless)
	A:	the frontal area (m^2)
	m:	vehicle mass (metric tons)

The basic concept of MOVES starts with the characterization of vehicle activity and the development of relationships between characterized vehicle activity and energy consumption, and between energy consumption and vehicle emission (Nam 2003). The U.S. EPA established a modal binning approach, developed using VSP, to characterize the relationship between vehicle activity and energy consumption. Originally, a total of 14 modal bins were developed based on different VSP ranges (U.S. EPA 2001a). This approach was revised in two different ways. U.S. EPA refined the VSP binning approach by the association of second-by-second speed, engine rpm, and acceleration rates, and the original 14 VSP binning approaches are revised with the combination of five different speed operating modes and redirected to a total of 37 VSP bins (Koupal et al. 2004). Researchers at North Carolina State University (NCSU) divided each bin into four strata representing two engine sizes and two odometer reading categories, and this approach was referred to as the "56-bin" approach. (U.S. EPA 2002b).

Another important conceptual model for MOVES was developed by NCSU in 2002 (Frey et al. 2002). Dr. Frey summarized the conceptual analytical methodology in the report "Recommended Strategy for On-Board Emission Data Analysis and Collection for the New Generation Model" (Frey et al. 2002). This method uses power demand estimate (P) as a variable on which emission rates can be based (Frey et al. 2002) as shown in Equation 3-4.

$$P = v \times a \qquad \text{(Equation 3-4)}$$

where: P : power demand (mph^2/sec)
v : vehicle speed (mph)
a : vehicle acceleration in (mph/s)

This method uses on-board emissions data where data are collected under real-world conditions to develop a modal emission model which can estimate emissions at different scales such as microscale, mesoscale, and macroscale. The philosophy is similar to MEASURE (Fomunung 2000), which first segregated the data into four modes based on suitable modal definitions, then developed an OLS regression model for each mode using explanatory variables selected by HTBR techniques. These explanatory variables include model year, humidity, temperature, altitude, grade, pressure, and power. Second and third powers of speed and acceleration were also included in the regression analysis.

3.3.4 HDDV-MEM

The researchers in Georgia Institute of Technology have developed a beta version of HDDV-MEM, which is based on vehicle technology groups, engine emission characteristics, and vehicle modal activity (Guensler et al. 2005). The HDDV-MEM first predicts second-by-second engine power demand as a function of on-road vehicle operating conditions and then applies brake-specific emission rates to these activity predictions. The HDDV-MEM consists of three modules: a vehicle activity module (with vehicle activity tracked by vehicle technology group), an engine power module, and an emission rate module. The model framework is illustrated in Figure 3-4.

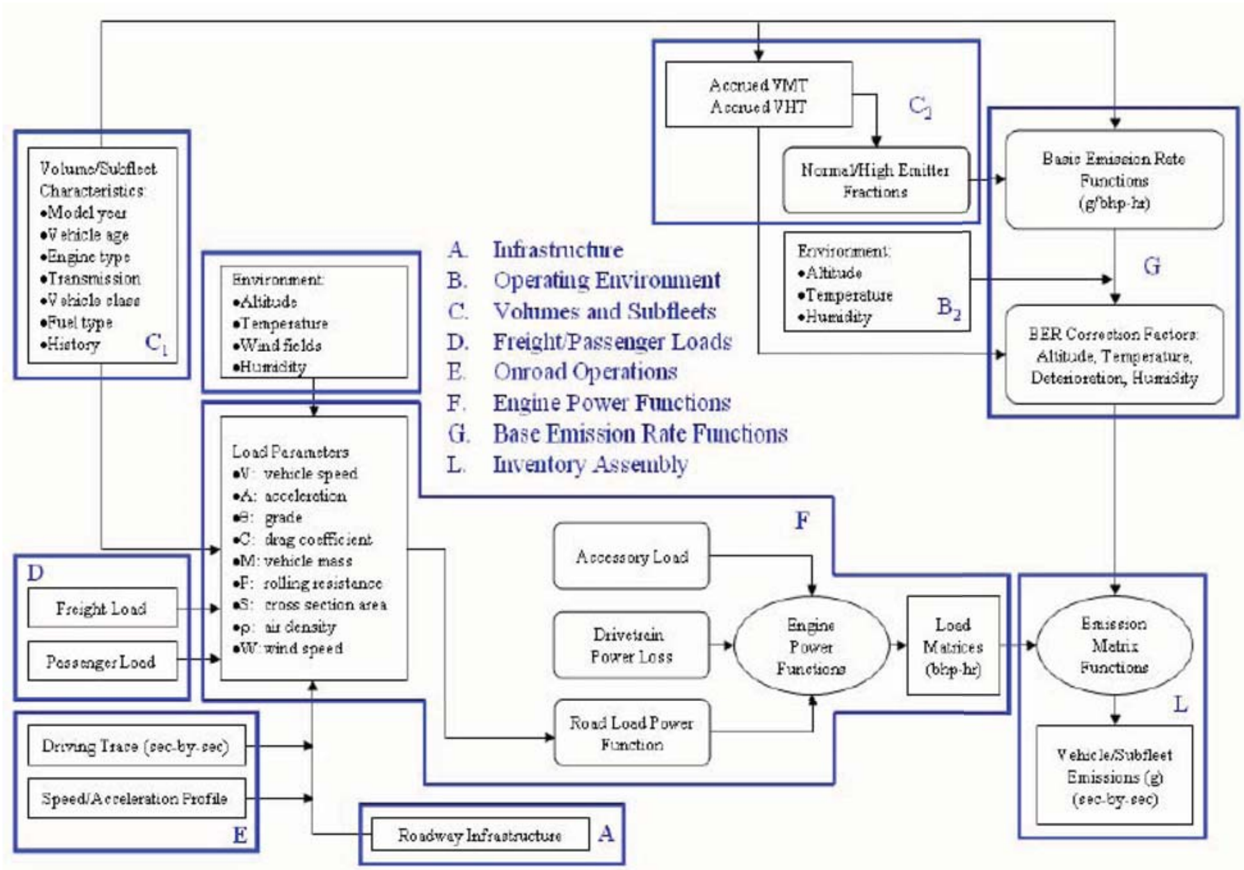


Figure 3-4 A Framework of Heavy-Duty Diesel Vehicle Modal Emission Model (Guensler et al. 2005)

3.3.4.1 Model Development Approaches

The HDDV-MEM modeling framework is designed for transportation infrastructure implementation on link-by-link basis. While the modeling routines are actually amenable to implementation on a vehicle-by-vehicle basis, the large number of vehicles operating on infrastructure links precludes practical application of the model in this manner. As such, the model framework capitalizes upon previous experience gained in development of the MEASURE modeling framework, in which vehicle technology groups were employed. A new heavy-duty vehicle visual classification scheme, which is an EPA and Federal Highway Administration (FHWA) hybrid vehicle classification scheme developed by Yoon et al. (Yoon et al. 2004b), classified vehicle technology groups by engine horsepower ratings, vehicles GVWR, vehicle configurations, and vehicle travel characteristics (Yoon 2005c). On the other hand, the MEASURE model employs load surrogates for the implementation of a light-duty modal modeling regime. This new modeling framework directly implements heavy-duty vehicle operating loads and uses these load predictions in the emission prediction process. An engine power module is designed for this task.

Emission rates are first established for various heavy-duty technology groups (engine and vehicle family, displacement, certification group, drivetrain, fuel delivery system, emission control system, etc.) based upon statistical analysis of standard engine dynamometer certification data, or on-road emission rate data when available (Wolf et al. 1998; Fomunung et al. 2000). The following subsets will discuss three main modules in the HDDV-MEM.

3.3.4.2 Vehicle Activity Module

The vehicle activity module provides hourly vehicle volumes for each vehicle technology group on each transportation link in the modeled transportation system. The annual average daily traffic (AADT) estimate for each road link is processed to yield vehicle-hours of operation per hour for each technology group (using truck percentages, VMT fraction by vehicle technology group, diesel fraction, hourly volume apportionment of daily travel, link length, and average vehicle speed) (Guensler et al. 2005; Yoon 2005c), as shown in Equation 3-5.

$$VA_{v,h,s|f} = (AADT_s \times (NL_s / TNL) \times HVF_{v,h} \times VF_v \times DF_v) \times (SL_s / AS_v) \qquad \text{(Equation 3-5)}$$

where:

- VA: the estimated vehicle activity (veh-hr/hr):
- v: the vehicle technology group
- h: the hour of day
- s: the transportation link
- f: the facility type for the link
- $AADT_s$: the annual average daily traffic for the link (number of vehicles)
- NL_s: the number of lanes in the specific link direction
- TNL: the total number of lanes on the link
- $HVF_{v,h}$: the hourly vehicle fraction
- VF_v: the VMT fraction for each vehicle technology group
- DF_v: the diesel vehicle fraction for each technology group
- SL_s: the link length (miles)
- AS_v: the link average speed of the technology group (mph)

To estimate on-road running emissions from each link, two sets of calculations are performed. On-road vehicle activity (vehicle-hr) for each hour is multiplied by engine power demand for observed link operations (positive tractive power demand plus auxiliary power demand), and then by baseline emission rates (g/bhp-hr). These calculations are processed separately for each speed/acceleration matrix cell (Yoon et al. 2005b). Emissions from motoring/idling activity are calculated by the determination of the vehicle-hours of motoring/idling activity on each link for each hour and the multiplication of the baseline idle emission rate (g/hr).

3.3.4.3 Engine Power Module

Internal combustion engines translate linear piston work (force through a distance) to a crankshaft, rotating the crankshaft and creating engine output torque (work performed in angular rotation). The crankshaft rotation speed (engine speed in revolutions per minute) is a function of engine combustion and physical design parameters (mean effective cylinder pressure, stroke length, connecting rod angle, etc.). The torque available at the crankshaft (engine output shaft) is less than the torque generated by the pistons, in that there are torque losses inside the engine associated with operating a variety of internal engine components. Torque is transferred from the engine output shaft to the driveshaft via the transmission (sometimes through a torque-converter, i.e., fluid coupling) and through a series of gears that allows the drive shaft to rotate at different speeds relative to engine crankshaft speed. The drive shaft rotation is then transferred to the drive axle via the rear differential. The ring and pinion gears in the rear differential translate the rotation of the drive shaft by 90 degrees from the drive shaft running along the vehicle to the drive axle that runs across the vehicle. Torque available at the drive axle is now delivered directly to the drive wheels. This process generates the tractive force used to overcome road friction, wind resistance, road grade (gravity), and other resistive forces, allowing the vehicle to accelerate on the roadway. Figure 3-5 illustrates the primary components of concern.

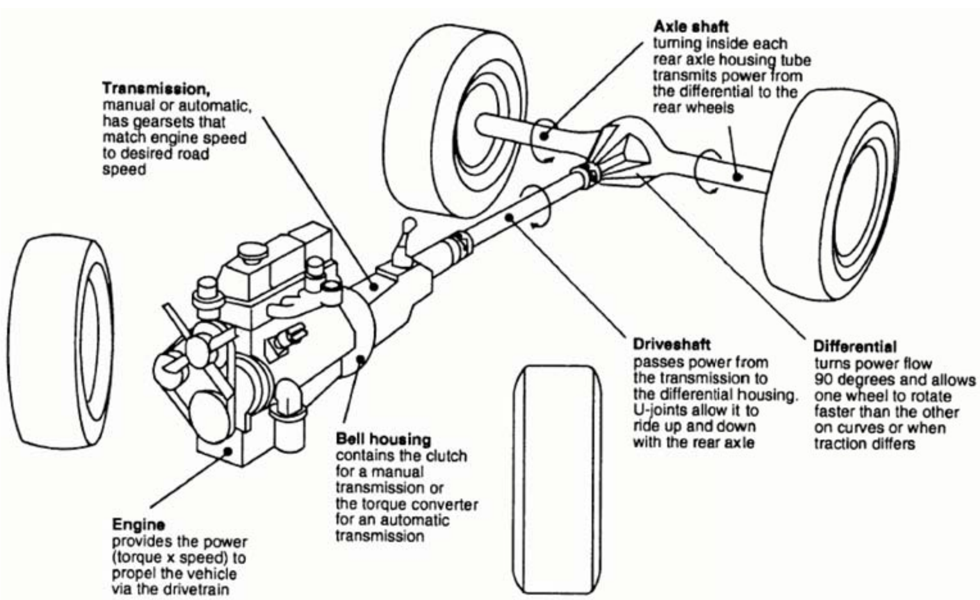


Figure 3-5 Primary Elements in the Drivetrain (Gillespie 1992)

The vehicle drivetrain (engine, torque converter, transmission, drive shaft, rear differential, axles, and wheels) is designed as a system to convert engine torque into useful tractive force at the wheel-to-pavement interface. When the tractive force is greater than the sum of forces acting against the vehicle, the vehicle accelerates in the direction of travel. Given that on-road speed/acceleration patterns for HDDVs can be observed (or empirically modeled), the modal modeling approach works backwards from observed speed and acceleration to estimate the tractive force (and power) that was available at the wheels to meet the observed conditions. Then, working backwards from tractive force, the model accounts for additional power losses that occurred between the engine and the wheels to predict the total brake-horsepower output of the engine. Force components that reduce available wheel torque and tractive force include:

- Aerodynamic drag, which depends on the frontal area, the drag coefficient, and the square of the vehicle speed;

- Tire rolling resistance, which is determined by the coefficient of rolling resistance, vehicle mass, and road grade (where the coefficient of rolling resistance is a function of tire construction and size; tire pressure; axle geometry, i.e., caster and camber; and whether the wheels are driven or towed);

- Grade load, which is determined by the roadway grade and vehicle mass; and

- Inertial load, which is determined by the vehicle's mass and acceleration.

The tractive force required at the interface between the tires and the road to overcome these resistive forces and provide vehicle acceleration can be described by (Gillespie 1992), as shown in Equation 3-6:

$$F_T = F_D + F_R + F_W + F_I + ma \quad \text{(Equation 3-6)}$$

where: F_T: the tractive force available at the wheels (lbf)
F_D: the force necessary to overcome aerodynamic drag (lbf)
F_R: the force required to overcome tire rolling res:tance (lbf)
F_W: the force required to overcome gravitational force (lbf)
F_I: the force required to overcome inertial loss (lbf)
m: the vehicle mass (lbm)
a: the vehicle acceleration (ft/sec^2)

Load prediction models could employ a wide variety of aerodynamic drag (Wolf-Heinrich 1998) and rolling resistance functional forms, some of which may be more appropriate for certain vehicle designs and at certain vehicle speeds. Note that vehicle mass is a critical parameter that must be included in the load-based modeling approach. Therefore, estimates of gross vehicle weight must be included in any transit (vehicle weight plus passenger loading) or heavy-duty truck (vehicle weight plus cargo payload) application. The following subsections describe each force in Equation 3-6, taken from Yoon et al. (Yoon et al. 2005a).

Aerodynamic Drag Force (F_D)

As a vehicle moves forward through the atmosphere, drag forces are created at the interface of the front of the vehicle and by the vacuum generated at the tail of the vehicle. The flow of the air around the vehicle creates a very complex set of forces providing both resistance to forward motion and vehicle lift. The net aerodynamic drag force is a function of air density, aerodynamic drag coefficient, vehicle frontal area, and effective vehicle velocity, as shown in Equation 3-7 (Yoon et al. 2005a).

$$F_D = (\frac{\rho}{2g}) \times C_d \times A_f \times V_e^{\,2} \qquad \text{(Equation 3-7)}$$

where:
- F_D: aerodynamic drag force
- ρ : the air density (lb/ft^3)
- g : the acceleration of gravity (32.2 ft/sec^2)
- C_d : the aerodynamic drag coefficient
- A_f : the vehicle frontal area (ft^2)
- V_e : the effective vehicle velocity (ft/sec)

Rolling Resistance Force (F_R)

Rolling resistance force is the sum of the forces required to overcome the combined friction resistance at the tires. Tires deform at their contact point with the ground as they roll along the roadway surface. Rolling resistance is caused by contact friction, the tires' resistance to deformation, aerodynamic drag at the tire, etc. The force required to overcome rolling resistance can be expressed with rolling resistance coefficient, vehicle weight, and road grade, as shown in Equation 3-8 (Yoon et al. 2005a).

$$F_R = C_r \times m \times g \times \cos(\theta) \qquad \text{(Equation 3-8)}$$

where:
- F_R: force required to overcome rolling resistance
- Cr: the rolling resistance coefficient
- θ : the road grade (degrees)
- m: vehicle mass in metric tons
- g: acceleration due to gravity

Gravitational Weight Force (F_W)

The gravitational force components account for the effect of gravity on vehicle weight when the vehicle is operating on a grade. The grade angle is positive on uphill grades (generating a positive resistance) and negative on downgrades (creating a negative resistance),as shown in Equation 3-9 (Yoon et al. 2005a).

$$F_W = m \times g \times \sin(\theta) \qquad \text{(Equation 3-9)}$$

where:
F_w: gravitational weight force
m: vehicle mass in metric tons
g: acceleration due to gravity
θ : the road grade (degrees)

Drivetrain Inertial Loss (F_I)

The engine, transmission, drive shaft, axles and wheels are all in rotation. The rotational speed of each component depends upon the transmission gear ratio, the final drive ratio, and the location of the component in the drive train (i.e., the total gear ratio between each component and the wheels). The rotational moment of inertia of the various drivetrain components constitutes a resistance to change in motion. The torque delivered by each rotating component to the next component in the power chain (engine to clutch/torque converter, clutch/torque converter to transmission, transmission to drive shaft, drive shaft to axle, axle to wheel) is reduced by the amount necessary to increase angular rotation of the spinning mass during vehicle acceleration. Given the torque loss at each component, the reduction in motive force available at the wheels due to inertial losses along the drivetrain can be modeled (Wolf-Heinrich 1998). This model term is most significant under low speed acceleration conditions, such as vehicle operation in truck and rail yards where vehicles are lugging heavy loads over short distances. However, as will be discussed later, significant new data will be required to incorporate the inertial loss effects into modal models, as shown in Equation 3-10 (Yoon et al. 2005a).

$$F_I = \frac{a \times I_{EFF}}{r^2} = \frac{a \times [(I_W + (G_d^2 \times I_D) + (G_t^2 \times G_d^2) \times (I_E + I_t)]}{r^2} \qquad \text{(Equation 3-10)}$$

where: a : the acceleration in the direction of vehicle motion (ft/sec2)
I_{EFF} : the effective moment of inertia (ft- lbf -sec2)

I_W : the rotational moment of inertia of the wheels and axles (ft-lbf -sec2)
I_D : the rotational moment of inertia of the drive shaft (ft-lbf -sec2)
I_T : the rotational moment of inertia of the transmission (ft-lbf -sec2)
I_E : the rotational moment of inertia of the engine (ft-lbf -sec2)
G_t : the gear ratio at the engine transmission
G_d : the gear ratio in the differential
r : wheel radius (ft)

Power Demand

Using the equations outlined above, the total engine power demand, which is the combination of tractive power and auxiliary power demands, can be expressed in Equation 3-11 (Yoon et al. 2005a):

$$P = [(\frac{V}{550}) \times (F_D + F_R + F_W + F_I + ma)] + AP \quad \text{(Equation 3-11)}$$

where P: total engine power demand
V : the vehicle speed (ft/s)
F_D: the force necessary to overcome aerodynamic drag (lbf)
F_R: the force required to overcome tire rolling res:tance (lbf)
F_W: the force required to overcome gravitational force (lbf)
F_I: the force required to overcome inertial loss (lbf)
m: the vehicle mass (lbm)
a: the vehicle acceleration (ft/sec2)
AP : the auxiliary power demand (bhp)
550 : the conversion factor to bhp

3.3.4.4 Emission Rate Module

The emission rate module provides work-related emission rates (g/bhp-hr) and idle emission rates (g/hr) for each technology group. The basic application of the HDDV-MEM incorporates a simple emission rate modeling approach. The predicted engine power demand (bhp) for each second of vehicle operation is multiplied by emission rates in gram/bhp-sec for a given bhp load. Technology groups (i.e., vehicles that perform similarly on the certification tests) are established based upon the engine and control system characteristics and each technology group is assigned a constant g/bhp-sec emission rate based upon regression tree and other statistical analysis of certification data. Under the assumption that testing cycles represent the typical modal activities undertaken by on-road activities, such emission rates are applied to on-road activity data. Given the large repository of certification data, detailed statistical analysis of the certification test results can be used to obtain applicable emission rates for these statistically derived vehicle technology groups. The data required for analysis must come from chassis dynamometer (the engine remains in the vehicle and the vehicle is tested on a heavy-duty treadmill) and on-road test programs in which second-by-second grams/second emission rate data have been collected concurrently with axle-hp loads.

At this moment, HDDV-MEM accepts EPA's baseline running emission rate data as work-related emission rates and EMFAC2002 idling emission rate test data as idle emission rates. Diesel vehicle registration fractions and annual mileage accumulation rates are employed to develop calendar year emission rates for each technology group. In the future, a constant emission rate need not be used as more refined testing data become available. Linear, polynomial, or generalized relationships can be established between gram/second emission rate and tractive horsepower (axle horsepower) and other variables. Sufficient testing data are required to establish statistically significant samples for each technology group.

3.3.4.5 Emission Outputs

HDDV-MEM outputs link-specific emissions in grams per hour (g/hr) for VOCs, CO, NO_X, and PM for each vehicle type. Toxic air contaminant emission rates (benzene, 1, 3-butadiene, formaldehyde, acetaldehyde, and acrolein) are also estimated in grams/hour for each vehicle type using the MOBILE6.2-modeled ratios of air toxics to VOC for each calendar year. HDDV-MEM provides not only hourly emissions, but also aggregated total daily emissions (in accordance with input command options). The structure of output files, which provide link-specific hourly emissions, can be directly incorporated with roadway network features in a GIS environment for use in interactive air quality analysis in various spatial scales, i.e., national, regional, and local scales (Guensler et al. 2005; Yoon 2005c).

CHAPTER 4

4. EMISSION DATASET DESCRIPTION AND POST-PROCESSING PROCEDURE

Using second-by-second data collected from on-road vehicles (Brown et al. 2001, Ensfield 2002), the research effort reported here developed models to predict emission rates as a function of on-road operating conditions that affect vehicle emissions. Such models should be robust and ensure that assumptions about the underlying distribution of the data are verified and that assumptions associated with applicable statistical methods are not violated. Due to the general lack of data available for development of heavy-duty vehicle modal emission rate models, this study focuses on development of an analytical methodology that is repeatable with different datasets collected across space and time. There are two second-by-second data sets in which emission rate and applicable load and vehicle activity data have been collected in parallel (Brown et al. 2001, Ensfield 2002). One database was a transit bus dataset, collected on diesel transit buses operated by Ann Arbor Transit Authority (AATA) in 2001 (Ensfield 2002), and another dataset was heavy HDV (HDV8B) dataset prepared by National Risk Management Research Laboratory (NRMRL) in 2001 (Brown et al. 2001). Each is summarized in the following sections.

4.1 Transit Bus Dataset

Transit bus emissions dataset was prepared by Sensors, Inc. (Ensfield 2002). Sensors, Inc. has supplied gas analyzers and portable emissions testing systems worldwide for over three decades. Their products, SEMTECH-G for gasoline powered vehicles, and SEMTECH-D for diesel powered vehicles, are commercially available for on-vehicle emission test applications. In October 2001, Sensors, Inc. conducted real-world, on-road emissions measurements of 15 heavy-duty transit buses for U.S. EPA (Ensfield 2002). Transit buses were provided by the AATA and all of them were New Flyer models with Detroit Diesel Series 50 engines. Table 4-1 summarizes the buses tested for U.S. EPA.

Table 4-1 Buses Tested for U.S. EPA (Ensfield 2002)

Bus #	Bus ID	Model Year	Odometer	Engine series	Displacement (liters)	Peak Torque (lb-ft)	Test Date
1	BUS360	1995	270476	SERIES 50 8047 GK40	8.5	890	10/25/2001
2	BUS361	1995	280484	SERIES 50 8047 GK38	8.5	890	10/25/2001
3	BUS363	1995	283708	SERIES 50 8047 GK37	8.5	890	10/24/2001
4	BUS364	1995	247379	SERIES 50 8047 GK42	8.5	890	10/24/2001
5	BUS372	1995	216278	SERIES 50 8047 GK41	8.5	890	10/26/2001
6	BUS375	1996	211438	SERIES 50 8047 GK39	8.5	890	10/25/2001
7	BUS377	1996	252253	SERIES 50 8047 GK36	8.5	890	10/24/2001
8	BUS379	1996	260594	SERIES 50 8047 GK35	8.5	890	10/23/2001
9	BUS380	1996	223471	SERIES 50 8047 GK28	8.5	890	10/23/2001
10	BUS381	1996	200459	SERIES 50 8047 GK29	8.5	890	10/22/2001
11	BUS382	1996	216502	SERIES 50 8047 GK30	8.5	890	10/17/2001
12	BUS383	1996	199188	SERIES 50 8047 GK31	8.5	890	10/19/2001
13	BUS384	1996	222245	SERIES 50 8047 GK32	8.5	890	10/17/2001
14	BUS385	1996	209470	SERIES 50 8047 GK33	8.5	890	10/18/2001
15	BUS386	1996	228770	SERIES 50 8047 GK34	8.5	890	10/19/2001

4.1.1 Data Collection Method

A total of 15 files were provided for the purpose of model development (Ensfield 2002). Each file represents data collected from different transit buses. Five of these buses were 1995 model year and the rest were 1996 model year. All of the bus test periods lasted approximately two hours. The buses operated along standard Ann Arbor bus routes and stopped at all regular stops although the buses did not board or discharge any passengers. The routes were mostly different for each test, and were selected for a wide variety of driving conditions. All of the bus routes for the test are shown in Figure 4-1.

Figure 4-1 Bus Routes Tested for U. S. EPA (Ensfield 2002).

Sensors, Inc. engineers performed the instrument setup and data collection for all the buses. Test equipment, SEMTECH-D analyzer, is shown in Figure 4-2. Because engine computer vehicle interface (SAE J1708) data were collected at 10 Hz, Sensors, Inc. engineers manually started and stopped data collections at approximately 30 minute intervals to keep file size manageable. A total of four trip files were generated per bus. Zero drift was checked between data collections. Then four files for each bus were combined into one file after post-processing. The time for each bus is thus sometimes not continuous. To derive other variables easily, like acceleration, and keep data manageable or other purposes, data for each bus were separated into trips based on continuous time. After this processing, there were 62 “trips” in the transit bus database.

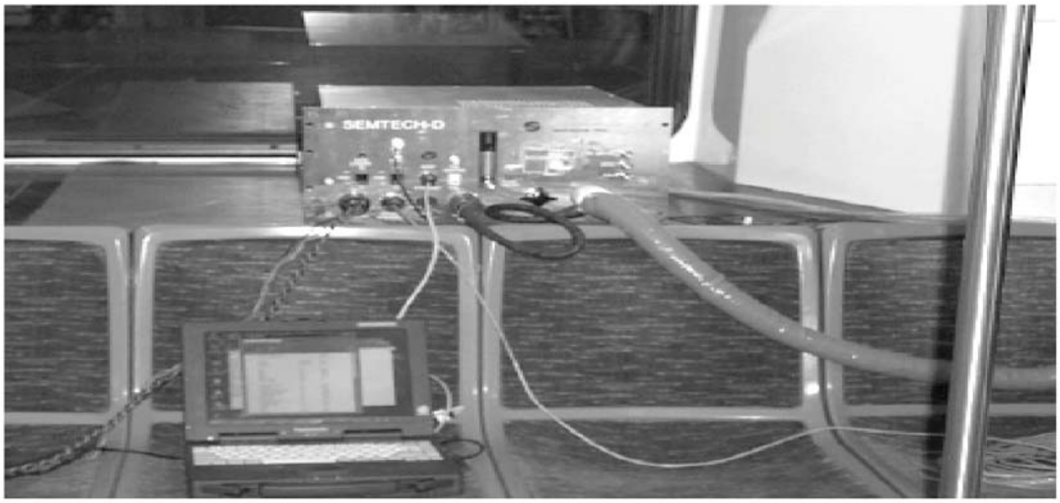


Figure 4-2 SEMTECH-D in Back of Bus (Ensfield 2002)

4.1.2 Transit Bus Data Parameters

Each of the 15 data files share the same format. The data fields included in each file are summarized in Table 4-2.

Table 4-2 Transit Bus Parameters Given by the U.S. EPA (Ensfield 2002)

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	License number; Engine size; Instrument configuration number
Roadway Characteristics	GPS Latitude (degree); GPS Longitude (degree); GPS Altitude (feet); Grade (%)
Onroad Load Parameters	Vehicle speed (mph); Engine speed (rpm); Torque (lb-ft); Engine power (bhp)
Engine Operating Parameters	Engine load (%); Throttle position (0 – 100%); Fuel volumetric flow rate (gal/s); Fuel specific gravity; Fuel mass flow rate (g/s); Calculated instantaneous fuel economy (mpg); Engine Oil temperature(deg F); Engine oil pressure (kPa); Engine warning lamp (Binary); Engine coolant temperature (deg F); Barometric pressure reported from ECM (kPa); Calculated exhaust flow rate (SCFM)
Environment Conditions	Ambient temperature (deg C); Ambient pressure (mbar); Ambient relative humidity (%); Ambient absolute humidity (grains/lb air)
Vehicle Emission	HC, CO, NO_x, CO2 emission (in ppm, g/sec, g/ke-fuel, g/bhp-hr units)

4.1.3 Sensors, Inc. Data Processing Procedure

It is helpful to understand how Sensors, Inc. processed the dataset after data collection This information is very important for data quality assurance and quality control. This section is adapted primarily from the Sensor's field data collection report (Ensfield 2002).

Data Synchronization: According to Sensor's report, the analytical instruments, vehicle interface, and global positioning system (GPS) equipment reported data individually to the SEMTECH data logger asynchronously and at differing rates, but with a timestamp at millisecond precision. The first step of the post-processing procedure is to eliminate the extra data by interpolating and synchronizing all the data to 1 Hz. With all the raw data synchronized to the same data rate, it is then time-aligned so that engine data corresponds to emissions data in real time.

Mass Emissions Calculations: Mass emissions (gram/second) are calculated by fuel flow method. With access to real-time, second-by-second fuel flow rates, a value for transient mass emissions is computed as shown by the equation below. Using NO as an example, NO mass emissions are calculated on a second-by-second basis (Ensfield 2002).

$$NO_{g/\sec} = NO_{fs} \times Fuelflow \qquad \text{Equation 4-1}$$

where $NO_{(g/sec)}$: NO emissions (grams/second)
NO_{fs} : NO emission rate (grams of NO per gram of fuel)
Fueflow : flow of fuel per unit time (grams per second).

Fuel specific emissions are the ratios of the mass of each pollutant to the fuel in the combusted air/fuel mixture. The mass fuel flow rate is converted from fuel volumetric flow rate using fuel specific gravity.

Brake Specific Emissions Calculations: Engine torque is first computed by applying the engine load parameter, which represents the ratio between current engine torque and maximum engine torque, to the engine lug curve (maximum torque curve). Engine horsepower is then converted from engine torque using engine speed data. Work (bhp-hr) is computed for each second of the test, and brake specific emissions are reported as the sum of the grams of pollutant emitted over the desired interval (one second) divided by the total work.

Vehicle Speed Validation: Vehicle speed is a critical parameter that influences the derived parameters, acceleration and emission rates. It is important for researchers to understand the method of measurement and data accuracy. Sensors, Inc. measured vehicle speed using two methods: vehicle Electronic Control Module (ECM) and Global Positioning System (GPS). Figure 4-3 shows the GPS vs. ECM comparison for Bus 380. The regression analysis shows that the ECM data are around 10% higher than the GPS data, according to Sensors report (Ensfield 2002). Sensors, Inc. researchers believe that this comparison suggests that GPS data may be more reliable for on-road testing. Buses of model year 1995 were equipped with an earlier version ECM that did not provide vehicle speed and GPS velocity data were used in place of the ECM data. Buses of model year 1996 were equipped with the current version ECM that can provide vehicle speed and vehicle speed was reported after validation with the GPS data. GPS data were within 1% accuracy based upon analysis of 10 miles of data (Ensfield 2002).

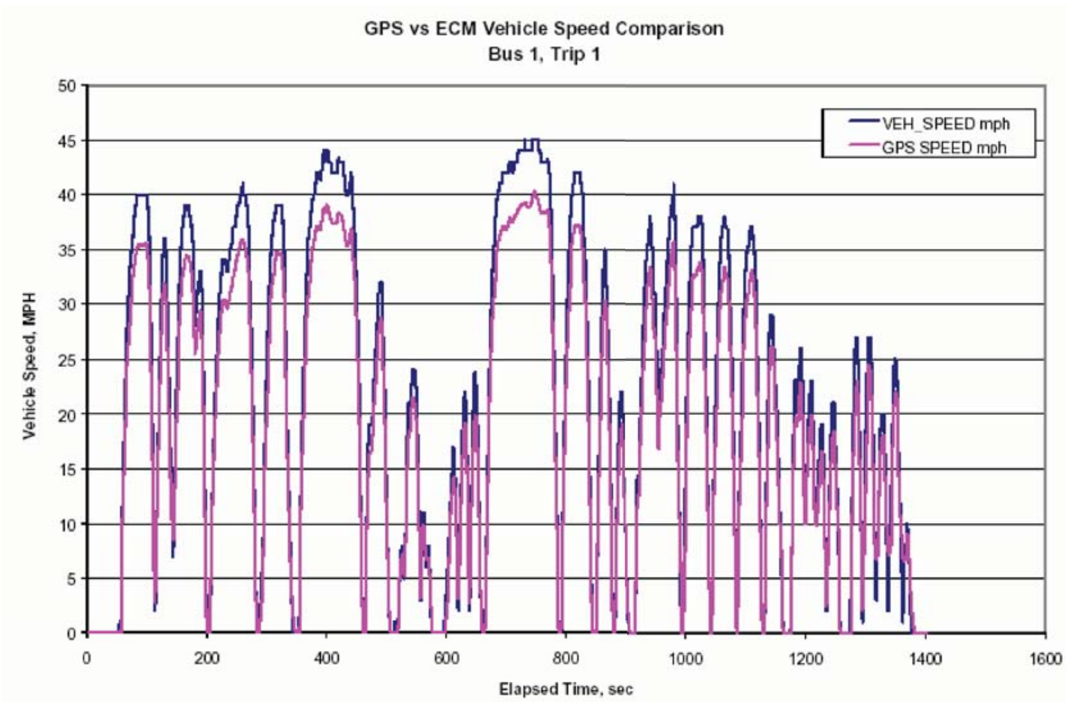


Figure 4-3 Bus 380 GPS vs. ECM Vehicle Speed (Ensfield 2002)

4.1.4 Data Quality Assurance/Quality Check

After understanding the manner in which Sensors, Inc. processed the reported data set, the data set for each bus was screened to check for errors or possible problems. Possible sources of errors associated with data collection should be considered before undertaking data analysis for the development of a model. The types of errors checked are listed below.

Loss of Data: Emission data are missing for some buses. For example, bus 382 had missing HC data for 343 seconds. Buses 361, 377 and 384 have similar problems. There might be several reasons for loss of data. Communication between instruments might be lost or a particular vehicle may have failed to report a particular variable. These records are removed from the test database and not employed in development of HC models because the instantaneous emission values will be recorded as zero, introducing significant bias to the result. Similarly, calculated fuel economy data are missing for some buses.

Erroneous ECM Data: There were some cases where certain engine parameters were well outside physical limits, and these erroneous ECM data were filtered out with pre-defined filter limits. The following filter limits (Ensfield 2002) were imposed on the rate of change of RPM, fuel flow, and vehicle speed data:

- Rate of change limit for RPM = 10,000 (RPM)/sec
- Rate of change limit for Fuel flow = 0.003 (gal/sec)/sec
- Rate of change limit for Vehicle speed = 21 (mph)/sec

According to Sensors, Inc. report, these filters remove the data outside the defined limits. The SEMTECH post-processor automatically interpolates between the remaining data, and produces results at 1Hz as before (Ensfield 2002). Because this procedure was finished by manually plotting the ECM parameters and computed mass results, all the buses' data were screened again to check any remaining data spikes for data quality assurance purposes. No such errors were identified for this kind of problem. But the modeler should keep in mind that data could be erroneous because "unreasonable" engine acceleration or deceleration was removed that could have been within reasonable absolute limits.

GPS Dropouts: There were a few instances when the GPS lost communication with the satellite for unknown reasons, and these erroneous GPS data were removed manually (Ensfield 2002). To guarantee data quality, the modeler screened all GPS data again to check any remaining erroneous cases. The principles for screening erroneous GPS data are based on the consistency between GPS data and engine parameters. The secondary screening identified that bus 360 data still contained some erroneous GPS data. The questionable area covers the beginning 434 seconds of the whole trip (see Figure 4-4). Their GPS data are shown as red in the left figure. The right figure illustrates the time series plot for checked area. Although GPS signals are reported as some fixed positions in the left figure while vehicle speed data are reported as zero in the right figure, engine speed and engine power in the right figure shows that bus 360 did move during that period. This error might due to GPS dropouts.

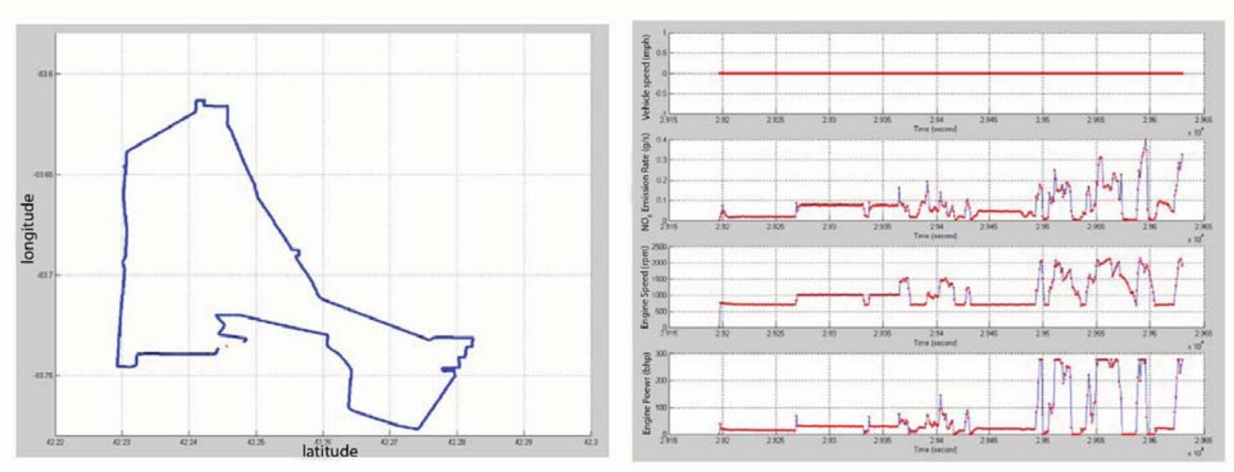


Figure 4-4 Example Check for Erroneous GPS Data for Bus 360 (Ensfield 2002)

Due to GPS dropouts, the GPS signals were reported as some fixed positions. At the same time, the vehicle speed might be reported as zero while other ECM data, such as engine speed and engine power, would show that the bus did move during that period. If the modeler fails to screen and remove such data, these data will be classified as idle mode. Further, these data will cause erroneous analysis result for idle mode. The modeler screened all buses manually and found that six buses had such problems (buses 360,361, 363, 364, 375 and 377). Usually, this type of error was prevalent during the beginning of the bus trip. All erroneous data were removed manually. The correction of the database to remove these erroneous data is critical to model development (initial models associated with development of idle and load-based emission rates were problematic until this database error was identified and corrected by the author).

Synchronization Errors: Data were checked for synchronization errors. An example plot of such a check is presented in Figure 4-5 where part of the trip for Bus 360 is used. The selected area covers about 200 seconds. Their GPS data are shown as the green/red part in the left figure. The figure on the right illustrates the time series plot for the area checked. The speed for red points in both figures is 0 mph. Although NOx correlates well to engine load and engine speed, vehicle speed doesn't correlate well to engine data and NOx emissions data. Bus 360 was equipped with an earlier version ECM that did not provide vehicle speed. GPS velocity data were used in place of the ECM data. According to Sensor's report, data synchronization was only done between emissions data and engine data, not for vehicle speed for emissions data (Ensfield 2002).

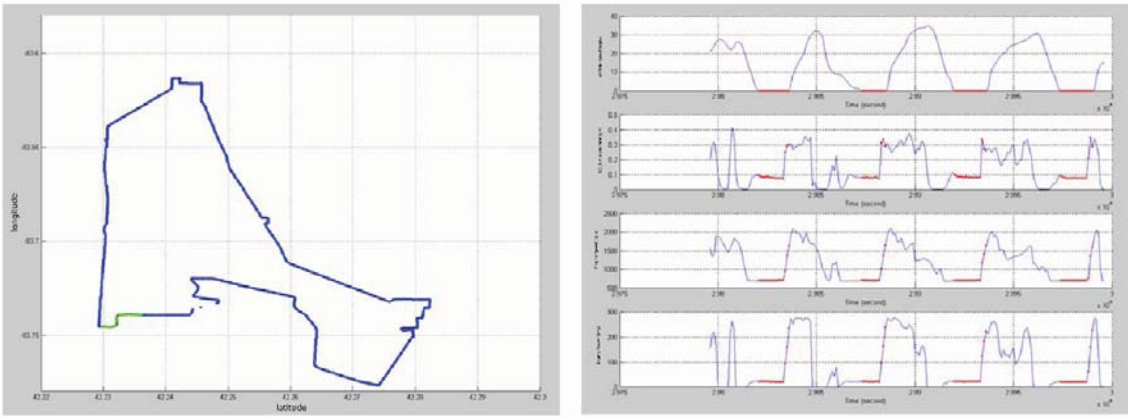


Figure 4-5 Example Check for Synchronization Errors for Bus 360

All bus data were checked for this type of error and such errors were identified in all of the test data for six buses (buses 360, 361, 363, 364, 375, 377). Coincidentally, these six buses had GPS dropout problems, too. From Frey's work (Frey and Zheng 2001), small errors in synchronization do not substantially impact estimate of total trip emissions. Such deviations will influence the estimate for micro-scale analysis. To choose the right delay time to remove the GPS data and vehicle speed data, the author compared the impacts of using a 2-second, 3-second, and 4-second delay. Figure 4-6 illustrates histograms of engine power for zero speed data based on three different proposed time delay options. A 3-second delay is chosen because engine power distribution for zero speed data based on a 3-second delay is more reasonable. Comparing to the 2-second delay results, zero speed data contain fewer data points with higher engine power (>150 brake horsepower) for 3-second delay. Meanwhile, zero speed data contain more data points with lower engine power (<20 brake horsepower) for a 3-second delay than 4-second delay time.

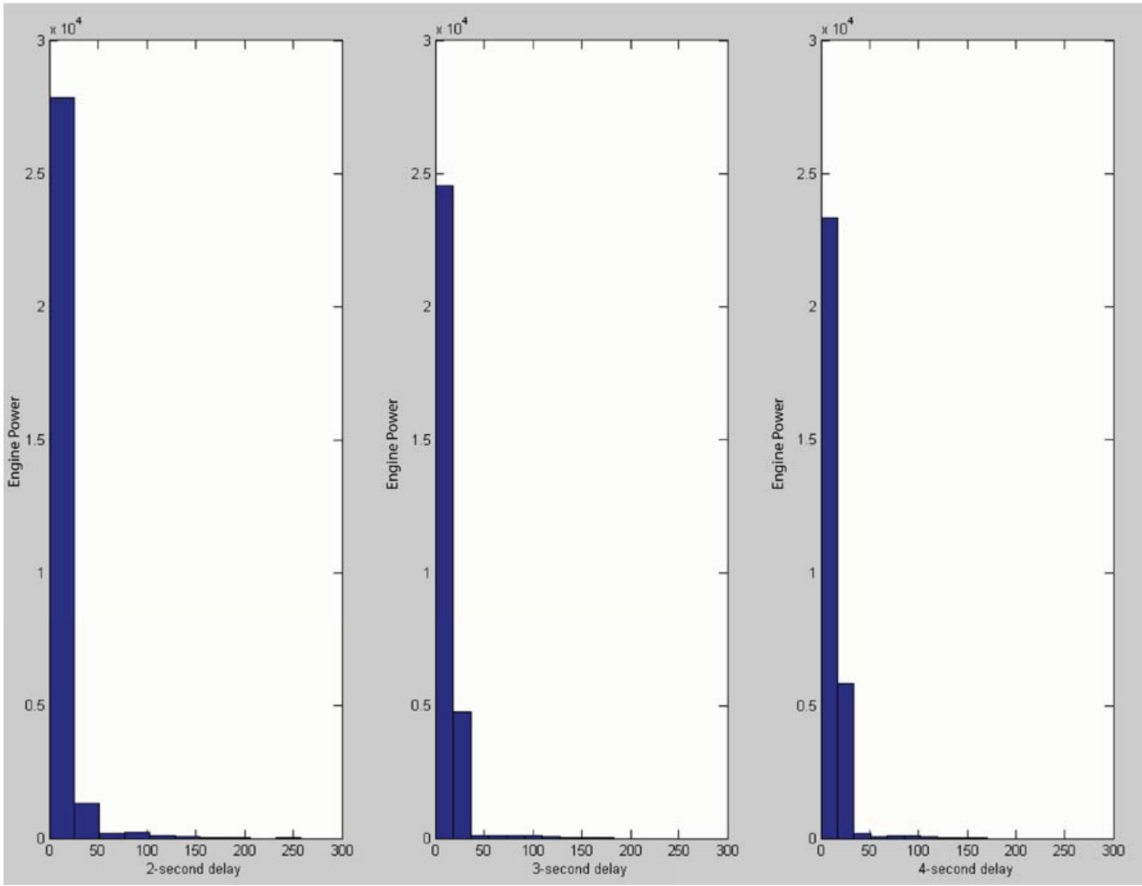


Figure 4-6 Histograms of Engine Power for Zero Speed Data Based on Three Different Time Delays

Road Grade Validation: According to Sensor's report, the GPS data were used for grade calculation. Combing the velocity at time t with the difference in altitude between time t and t-1 second, the instantaneous grade is computed as shown in Equation 4-2 (Ensfield 2002).

$$\text{Grade}_t = \frac{\text{velocity}_t}{\text{altitude}_t\text{-altitude}_{t-1}} \qquad \text{Equation 4-2}$$

where grade_t : Road grade at time t
t : time, t or t-1 second
velocity_t : vehicle speed in feet per second at time t
altitude : altitude in feet at time t or t-1

The calculation formula can generate significant errors given the uncertainty in the GPS position, particularly at low speeds where there is less of a differential in distance over the one-second interval (Ensfield 2002). In the real world, the maximum recommended grade for use in road design depends upon the type of facility, the terrain on which it is built, and the design speed. Figure 4-7 is directly cited from Traffic Engineering (Roess et al. 2004) to present a general overview of usual practice. Roess et al. (2004) indicated that these criteria represent a balance between the operating comfort of motorists and passengers and the practical constrains of design and construction in more severe terrains.

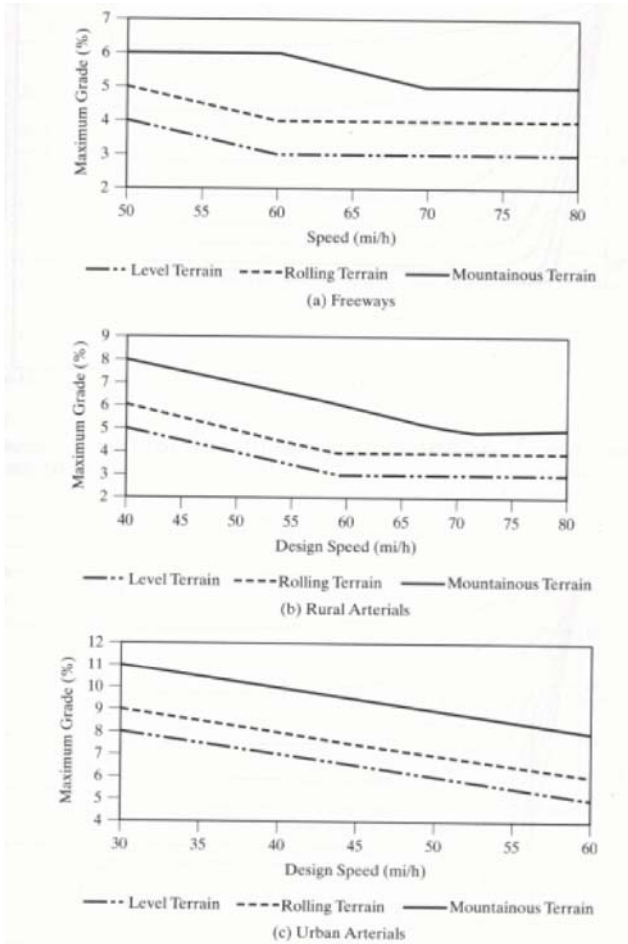


Figure 4-7 General Criteria for Maximum Grades (Roess et al. 2004)

The modeler screened the grade data in the database and found that 0.42% of the data have higher grade (> 10%). Meanwhile, 2% of the road grade data have higher rate of change (> 5%). This means some road grade data are dubious or erroneous. Considering Sensors, Inc. recommendations, road grade data would only be used as reference, and would not be used directly in model development.

4.1.5 Database Formation

The data dictionaries of the source files were reviewed for parameter content. Not all variables reported will be included in explanatory analysis. A standard file structure was designed to accommodate the available format. Emissions rate data with units of grams/second were selected to develop the proposed emission rate model. Because volumetric fuel rate, fuel specific gravity, and fuel mass flow rate are used to calculate mass emissions (g/s), these variables will be excluded in further analysis. Similarly, because percent engine load, engine torque, and engine speed are used to calculate engine power (brake horsepower), only engine power (bhp) is selected to represent power related variables. Exhaust flow rate is excluded because it is back-computed from the mass emissions generated with the fuel flow method. Fuel economy is excluded because it is 30 second moving average data and computed for a test period by summing the fuel consumed and dividing by the distance traveled. Because GPS data were used for grade calculation and road grade data would only be used as reference, a dummy variable was created to represent different road grade ranges.

At the same time, variables that might be helpful in explaining variability in vehicle emissions were included in the proposed file structure although they were not provided in the original dataset. These variables include model year, odometer reading, and acceleration. Acceleration data were derived from speed data using central difference method. Table 4-3 summarizes the parameter list for explanatory analysis.

Table 4-3 List of Parameters Used in Explanatory Analysis for Transit Bus

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	License number; Model year; Odometer reading; Engine size; Instrument configuration number
Roadway Characteristics	Dummy variable for road grade range
Onroad Load Parameters	Engine power (bhp); Vehicle speed (mph); Acceleration (mph/s)
Engine Operating Parameters	Throttle position (0 – 100%); Engine oil temperature (deg F); Engine oil pressure (kPa); Engine warning lamp (Binary); Engine coolant temperature (deg F); Barometric pressure reported from ECM (kPa)
Environmental Conditions	Ambient temperature (deg C); Ambient pressure (mbar); Ambient relative humidity (%); Ambient absolute humidity (grains/lb air)
Vehicle Emissions	HC, CO, NOx emission (in g/sec)

4.1.6 Data Summary

After the post-processing procedure was completed, the summary of the emissions and activity data as well as environmental and roadway characteristics is given in Table 4-4.

Table 4-4 Summary of Transit Bus Database

Bus ID	360	361	363	364	372	375	377	379	380	381	382	383	384	385	386
Numbers of Seconds of Data	7606	5153	7623	5284	5275	7323	7809	7880	8006	7282	3136	7943	8453	8423	10339
Vehicle Operation															
Average Speed (mph)	11.116	25.804	14.626	19.046	21.45	16.814	12.518	15.118	13.035	16.335	19.947	18.253	18.262	16.559	17.319
Average Engine Power (bhp)	71.952	87.536	65.822	79.599	72.395	86.307	78.121	84.82	72.987	65.724	85.224	67.249	64.199	62.512	62.979
Emission Data															
Average CO (g/s)	0.029652	0.018965	0.022419	0.020627	0.016582	0.031844	0.028571	0.030731	0.052504	0.034294	0.052822	0.026207	0.036183	0.023527	0.047062
Average Nox (g/s)	0.11049	0.1484	0.066047	0.12341	0.087625	0.13697	0.074597	0.10658	0.10393	0.090166	0.14089	0.11873	0.10457	0.095998	0.10635
Average HC (g/s)	0.001838	0.001304	0.000239	0.003492	0.002371	0.001377	0.000557	0.001807	0.001073	0.000609	0.00132	0.001803	0.00137	0.001698	0.00147
Environmental Characteristics															
Average Ambient Temperature (deg C)	20.358	16.666	25.623	20.358	21.375	17.5	26.012	23.788	23.648	22.465	21.746	21.282	18.17	21.842	20.389
Average Ambient Pressure (mbar)	977.16	971.08	965.69	985.58	982.05	977.52	973.08	974.27	973.22	987.82	994.71	983.55	992.7	991.34	985.65
Average Humidity (grains/ (lb air))	24.512	26.745	88.396	33.227	32.494	24.394	70.653	70.818	67.525	46.016	27.868	44.646	22.494	29.766	37.239

4.2 Heavy-duty Vehicle Dataset

The heavy-duty vehicle emission dataset is prepared by the U.S. EPA N ational Risk Management Research Laboratory (NRMRL) (U.S. EPA 2001b). EPA's Onroad Diesel Emissions Characterization (ODEC) facility has been collecting real-world gaseous emissions data for many years (U.S. EPA 2001c). The on-road facility incorporated a 1990 Kenworth T800 tractor-trailer as its test vehicle to collect this database. When this truck was purchased, it had already logged over 900,000 miles and was due for an overhaul of its Detroit Diesel Series 60 engine. The vehicle was tested prior to having this work done and after the overhaul. NRMRL collected the test data for U.S. EPA from 1999 to 2000 and included all the results and findings in a report titled: "Heavy Duty Diesel Fine Particulate Matter Emissions: Development and Application of On-Road Measurement Capabilities" (U.S. EPA 2001c).

4.2.1 Data Collection Method

The general capabilities of the ODEC facility are shown in Figure 4-8. The facility is designed to collect data while traveling along the public roadways using a 1990 Kenworth T800 tractor-trailer. This truck was tested using two types of tests. During 'parametric' testing, the truck systematically follows a test matrix representing the full range of load, grade, speed and acceleration conditions. During 'highway' testing, the truck travels along an interstate highway with no specific agenda other than covering the distance safely and efficiently; speed and acceleration vary randomly with grade, speed limit, and traffic effects. Tables 4-5 and 4-6 summarize the tests finished by NRMRL for U.S. EPA.

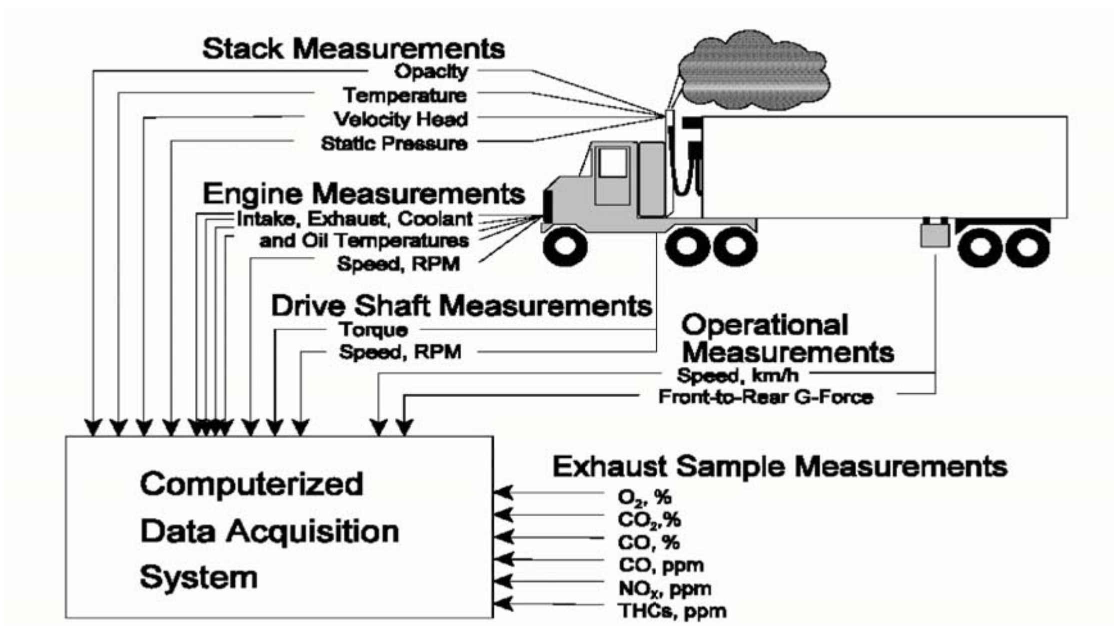


Figure 4-8 Onroad Diesel Emissions Characterization Facility (U.S. EPA 2001c)

Table 4-5 Onroad Tests Conducted with Pre-Rebuild Engine

Test ID	Load lb GCW	Grade(s) %	Comments
3F00V	79280	Zero	Constant Speed Testing
3F00C	79280	Zero	Cost Down & Acceleration
3F00A	79280	Zero	Governed Acceleration & Short-shift Acceleration
3H00V	61060	Zero	Constant Speed Testing
3H00C	61060	Zero	Cost Down & Acceleration
3H00A	61060	Zero	Governed Acceleration & Short-shift Acceleration
3E00V	42840	Zero	Constant Speed Testing
3E00C	42840	Zero	Cost Down & Acceleration
3E00A	42840	Zero	Governed Acceleration & Short-shift Acceleration
3F0GA	79280	Zero	Governed Acceleration
3F0SA	79280	Zero	Short-shift Acceleration
3F0V	79280	Zero	Constant Speed Testing
3H0GA	61060	Zero	Governed Acceleration
3H0SA	61060	Zero	Short-shift Acceleration
3H0V	61060	Zero	Constant Speed Testing
3E0GA	42840	Zero	Governed Acceleration
3E0SA	42840	Zero	Short-shift Acceleration
3E0V	42840	Zero	Constant Speed Testing
3F3&6	79280	3.1, 6.0	Uphill Grade Tests
3H3&6	61060	3.1, 6.0	Uphill Grade Tests
3E3&6	42840	3.1, 6.0	Uphill Grade Tests
3F-SEQ	79280	Zero	Dyno Sequence Simulations
3DRI	79280	Various	Open Highway Tests – Tunnel
3FIL	61060	Various	Open Highway Tests – Filters
3DIOX*	61060	Various	Open Highway Tests – Dioxin

*Note: These tests are not available.

Table 4-6 Onroad Tests Conducted with Post-Rebuild Engine

Test ID	Load lb GCW	Grade(s) %	Comments
5F0V	74000	Zero	Constant Speed Testing
5F0C*	74000	Zero	Cost Down & Acceleration
5F0A*	74000	Zero	Governed Acceleration & Short-shift Acceleration
5H0V	61440	Zero	Constant Speed Testing
5H0C*	61440	Zero	Cost Down & Acceleration
5H0A*	61440	Zero	Governed Acceleration & Short-shift Acceleration
5E0V	42600	Zero	Constant Speed Testing
5E0C*	42600	Zero	Cost Down & Acceleration
5E0A*	42600	Zero	Governed Acceleration & Short-shift Acceleration
5F3&6	74000	3.1, 6.0	Uphill Grade Tests
5H3&6	61440	3.1, 6.0	Uphill Grade Tests
5E3&6	42600	3.1, 6.0	Uphill Grade Tests
5F-SEQ*	74000	Zero	Dyno Sequence Simulations
5Plume	61440	Various	Open Highway Tests – Plume
5NOxB*	61440	Various	Open Highway Tests – Burst
5DIOX*	61440	Various	Open Highway Tests – Dioxin

*Note: These test results are not available.

4.2.2 Heavy-duty Vehicle Data Parameters

A total of 42 files were collected for the pre-rebuild engine and a total of 38 file collected for the post-rebuild engine. Each file represents data collected for a different engine and test. Preliminary analysis of individual files indicated that the format of files was same for all available files. The data fields included in each file are summarized in Table 4-7 below.

Table 4-7 List of Parameters Given in Heavy-duty Vehicle Dataset Provided by U.S. EPA

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	Vehicle make/model; Model year; Engine type; Engine Rating; Vehicle maintenance history
Onroad Load Parameters	Truck load weight (lb); Vehicle speed (mph); Measured engine power (bhp)
Engine Operating Parameters	Engine speed (RPM); Shaft volts; Torque volts; Fuel H/C ratio; Fuel factor; Engine intake air temperature (deg F); Engine exhaust air temperature (deg °F); Engine coolant temperature (deg °F); Engine oil temperature (deg °F)
Environment Conditions	Barometric pressure (inches Hg); Ambient humidity (%)
Vehicle Emissions	CO, NO_x, and HC emission (in ppm, g/hr, g/kg fuel and g/hp-hr units)

4.2.3 Data Quality Assurance/Quality Control Check

Although a total of 80 tests were finished for that project, preliminary screening found that there were some test files missing from the data DVD provided by U.S. EPA to the researchers. The missing test files include: 3DIOX, 5E0C, 5H0C, 5F0C, 5F-SEQ, 5NOxB, and 5DIOX. For quality assurance purposes, the available data files were screened to check for errors or possible problems. Possible sources of errors for data collection should be considered before developing the model. The types of errors checked are listed below.

Loss of Data: Measured horsepower (engine power) and emission data were missing for some tests. Tests 3F-SEQ, 3FIL1, 3FIL2, and 3FIL3 had no measured horsepower data for the entire test. These test files couldn't be included in emission model development. In addition, tests 3E00A, 3E00C, 3E00V, 3F0GA, 3F0SA, 3F0V, 3H0SA, 3FIL4, 3FIL5, 3FIL7, 3FIL8, 3FIL9, 3FIL10, and 5H0V had no HC emission data. This problem will be fixed by removing these tests for HC emission model development. Test 3H0SA also had no CO emission data and this problem will be treated by removing this test for CO emission model development.

Duplicated Records: A notable issue was duplicate records with different emission values for same time in some test files. After communicating with Mr. Brown who prepared this dataset for EPA, the reason was identified: the data were recorded at rates as high as 10 Hz to improve the resolution of the data. To keep consistent with other test files, these data were post-processed as one data point for each second.

Erroneous Load Data: The "measured horsepower" field is engine power data calculated from measurement of the drive shaft torque and rotational speed. Results from the literature review show that engine power is a major explanatory variable of possible erroneous load data. This variable was screened to check for errors or possible problems. An example of a check of measured horsepower is given in Figure 4-9. The observed relationship between measured horsepower and engine speed is to some extent a relationship between vehicle speed and engine speed which can be found in “Fundamentals of Vehicle Dynamics” (Gillespie 1992). At a given gear ratio, the relationship between engine speed and road speed is to some extent a linear relationship. The geometric progression in the left figure reflects the choices made in selection of transmission gear ratios. The right figure shows a problematic linear relationship between measured horsepower and vehicle speed. Essentially, the right figure appears to show no gear changes as vehicle speed increases, indicating that measured horsepower has been calculated incorrectly for this test. Such problems exist in the series of tests 3DRI and test 5Plume. These test files were removed from emission model development.

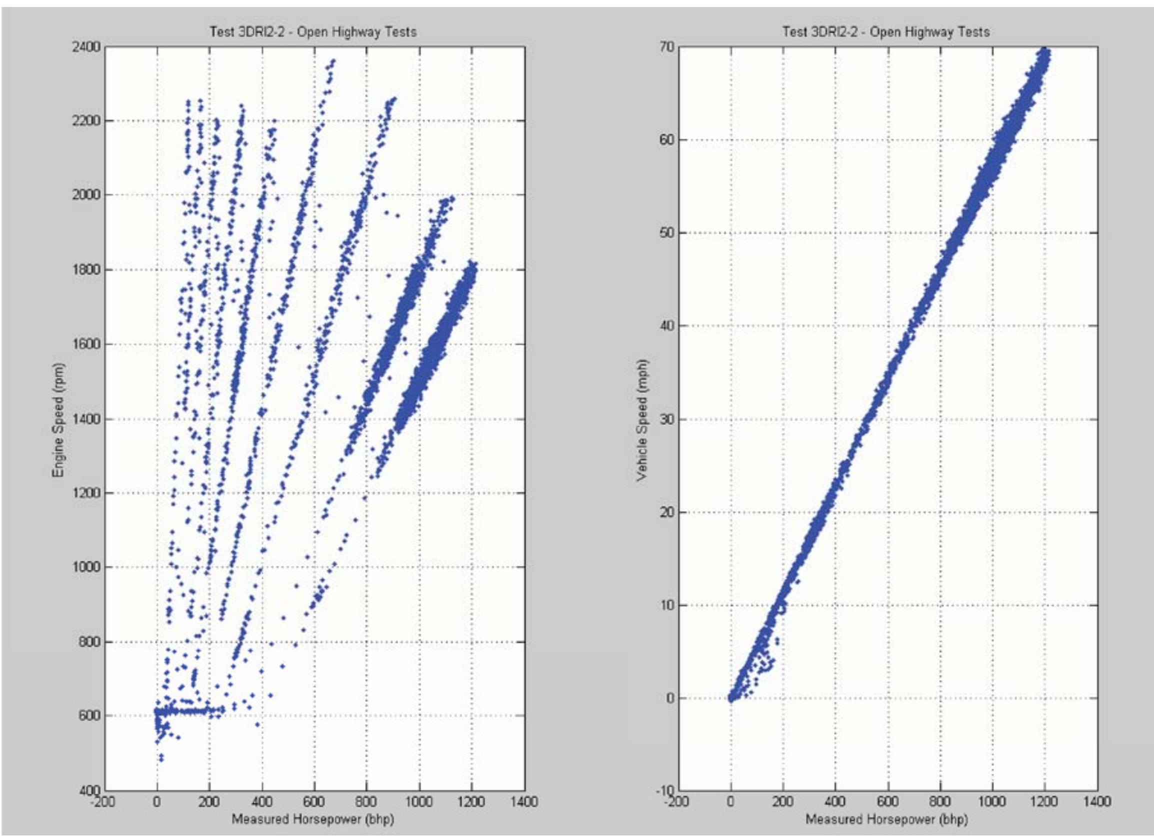


Figure 4-9 Example Check for Erroneous Measured Horsepower for Test 3DRI2-2

Vehicle Speed Validation: The author reviewed NRMRL’s report (U.S. EPA 2001c) related to vehicle speed validation. Vehicle speed data were measured with a Datron LS1 optical speed sensor. The product literature specifies an accuracy of +/- 0.2% and a reproducibility of +/- 0.1% over the measurement range of 0.5 to 400 kph. Figure 4-10 from NRMRL’s report correlates the speed measurement to a drive shaft speed sensor that was scaled using a National Institute of Standards and Technology (NIST)-traceable frequency source. The outliers at the low-speed indicated when the truck was turning (the tractor and the trailer-mounted speed sensor traveled less distance than the tractor does during turns). Notwithstanding these points, the correlation is a good indication of speed measurement precision.

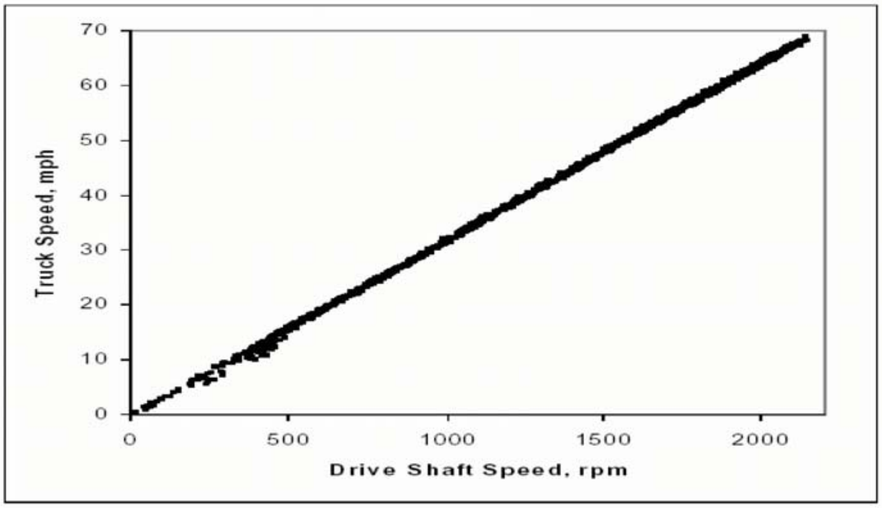


Figure 4-10 Vehicle Speed Correlation (U.S. EPA 2001c)

At the same time, NRMRL provided Figure 4-11 (U.S. EPA 2001c) to show the precision for four ranges of vehicle speed, along with similar estimates of accuracy. This figure will help researchers deal with speed measurement noise in the future.

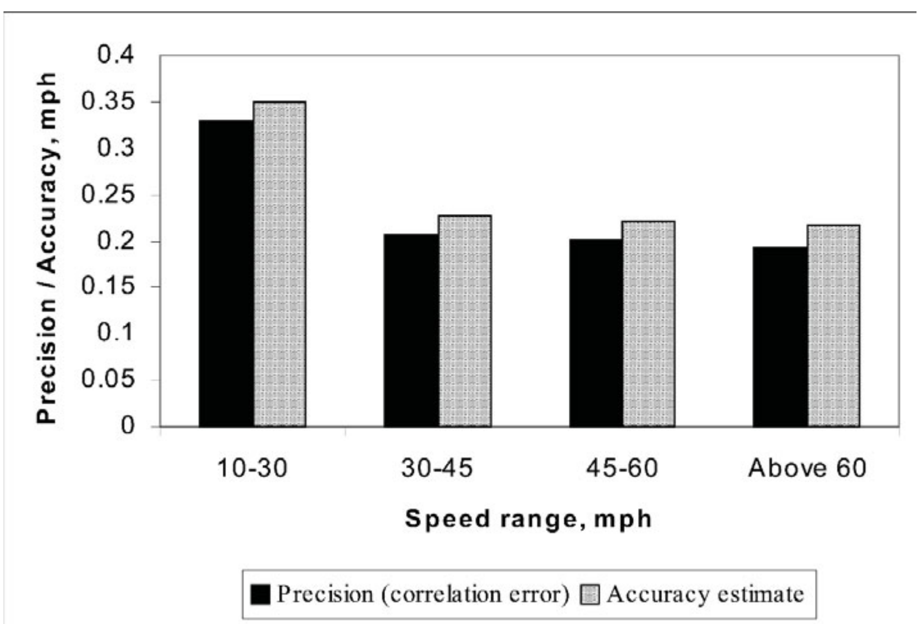


Figure 4-11 Vehicle Speed Error for Different Speed Ranges (U.S. EPA 2001c)

4.2.4 Database Formation

The data dictionaries of the source files were reviewed for parameter content (Table 4-8). Not all variables reported are included in explanatory analysis. A standard file structure was designed to accommodate the available format. Emissions data with units of gram/second were selected to develop the proposed emission model. All variables used to calculate mass emissions were excluded in further analysis. Similarly, because the "measured horsepower" field is calculated from measurements of drive shaft torque and rotational speed, only "measured horsepower" is used to represent power related variables. At the same time, variables like acceleration that might be helpful in explaining variability in vehicle emissions were included in the proposed file structure although they were not provided in the original dataset. Acceleration data were derived from speed data using the central difference method.

Table 4-8 List of Parameters Used in Explanatory Analysis for HDDV

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	Vehicle make/model; Model year; Engine type; Engine rating; Vehicle maintenance history
Onroad Load Parameters	Truck load weight (lb); Vehicle speed (mph); Acceleration (mph/s); Measured engine power (bhp)
Engine Operating Parameters	Engine intake air temperature (deg F); Engine exhaust air temperature (deg F); Engine coolant temperature (deg F); Engine oil temperature (deg F)
Environment Conditions	Barometric pressure (Hg), Ambient moisture (%)
Vehicle Emissions	CO, NO_x, and HC emission (in g/s units)

4.2.5 Data Summary

After the post-processing procedure was completed, a summary of the emissions and activity data as well as environmental and roadway characteristics is given in Table 4-9.

Table 4-9 Summary of Heavy-Duty Vehicle Data U.S. EPA 2001c).

Test ID	Number of Seconds of Data	Vehicle Operation		Emission Data			Environment Characteristics	
		Average Speed (mph)	Average Engine Power (bhp)	Average CO (g/s)	Average NO_x (g/s)	Average HC (g/s)	Barometric Pressure (Hg)	Ambient Moisture (%)
3F00V	4430	43.55	163.10	0.11633	0.27983	0.001442	28.273	1.6874
3F00C	7991	36.49	323.79	0.08200	0.19566	0.001166	28.272	1.6874
3F00A	1904	43.55	475.12	0.17476	0.34262	0.001471	28.272	1.6874
3H00V	3718	43.66	130.99	0.08386	0.22701	0.001429	28.273	1.6874
3H00C	7593	39.43	112.50	0.07456	0.17866	0.001414	28.272	1.6874
3H00A	1959	48.04	218.50	0.20521	0.32078	0.001751	30.423	1.3573
3E00V	3863	41.41	123.42	0.10896	0.21157	NA	28.273	1.6874
3E00C	7962	39.31	104.95	0.07489	0.14908	NA	28.272	1.6874
3E00A	1810	50.15	197.07	0.22324	0.26108	NA	30.137	1.9020
3F0GA	577	35.93	302.14	0.23114	0.41269	NA	29.995	0.4685
3F0SA	792	36.26	287.45	0.25140	0.37947	NA	29.995	0.4685
3F0V	3635	41.65	152.23	0.14879	0.28413	NA	29.995	0.4685
3H0GA	594	33.81	253.63	0.30036	0.48494	0.002159	29.690	1.6059
3H0SA	707	34.27	223.73	NA	0.32498	NA	29.690	1.6059
3H0V	3331	41.53	143.38	0.08892	0.27712	0.002436	28.020	0.4742
3E0GA	421	32.91	233.93	0.37978	0.30728	0.000589	29.976	0.5812
3E0SA	571	31.99	180.73	0.23652	0.33325	0.003042	29.976	0.5812
3E0V	3395	42.64	103.63	0.08879	0.25745	0.002805	29.976	0.5812
3F3&6	8629	36.59	131.00	0.14409	0.31374	0.001426	28.282	1.2520
3H3&6	10573	43.13	107.06	0.16769	0.27507	0.001753	28.273	1.6874
3E3&6	9825	44.74	121.69	0.16617	0.23913	0.001839	28.250	1.5716
3FIL4	12456	66.54	152.91	0.06994	0.29925	NA	29.238	0.3886
3FIL5	13738	58.76	129.99	0.06354	0.22315	NA	29.238	0.3886
3FIL6	6415	66.94	130.11	0.06273	0.20833	0.001409	29.238	0.3886
3FIL7	10678	62.76	164.82	0.07042	0.28353	NA	29.854	0.1480
3FIL8	12248	64.70	147.26	0.06688	0.26035	NA	29.773	0.1484
3FIL9	11956	65.62	153.44	0.06551	0.20905	NA	29.418	0.1502
3FIL10	12367	63.71	167.73	0.07481	0.35788	NA	30.132	0.1466
5F0V	4895	32.87	96.09	0.10716	0.23558	0.002828	30.101	0.5761
5H0V	4091	42.36	126.14	0.12564	0.30933	NA	30.179	0.6091
5E0V	4407	42.60	105.84	0.10681	0.29045	0.002894	30.278	0.8601
5F3&6a	6971	36.24	147.99	0.13716	0.31607	0.003111	28.004	0.9070
5F3&6b	5058	38.69	133.54	0.14044	0.30661	0.001924	28.009	0.8862
5H3&6a	6919	39.74	133.01	0.12723	0.28763	0.002397	28.024	0.8138
5H3&6b	6951	39.44	148.26	0.15400	0.32910	0.002807	28.014	1.2149
5E3&6	10807	46.01	124.07	0.13981	0.27674	0.002827	28.024	1.0131

CHAPTER 5

5. METHODOLOGICAL APPROACH

The following chapter lays the theoretical foundation of the conceptual framework of model development. This chapter outlines the statistical methods, addresses issues that arise in statistical modeling, and presents the solutions that are employed to address these issues. This chapter will serve as a guide or "road map" for the underlying methodology of the model development process.

5.1 Modeling Goal and Objectives

The goal of this research is to provide emission rate models that fill the gap between the existing models and ideal models for predicting emissions of NO_x, CO, and HC from heavy-duty diesel vehicles. Problems in existing models, like EPA's MOBILE series and CARB's EMFAC series of models, have been highlighted in previous chapters. U.S. EPA is currently developing a new set of modeling tools for the estimation of emissions produced by on-road and off-road mobile sources. MOVES, a new model under development by EPA's OTAQ, is a modeling system designed to better predict emissions from on-road operations. The philosophy behind MOVES is the development of a model that is as directly data-driven as possible, meaning that emission rates are developed from second-by-second or binned data.

Using second-by-second data collected from on-road vehicles, this research effort will develop models that predict emissions as a function of on-road variables known to affect vehicle emissions. The model should be robust and ensure that assumptions about the underlying distribution of the data are verified and the properties of parameter estimates are not violated. With limited available data, this study focuses on development of an analytical methodology that is repeatable with a different data set from across space and across time. As more data become available, the proposed model will need to be re-estimated to ensure that the model is transferable across additional HDV engine types, operating conditions, environmental conditions, and even perhaps geographical regions.

5.2 Statistical Method

The purpose of statistical modeling was to determine which explanatory variables significantly influence vehicle emissions so that the data can be stratified by those variables and a corresponding regression relationship can be developed. For many statistical problems there are several possible solutions. In comparing the means of two small groups, for instance, we could use a *t* test, a *t* test with a transformation, a Mann-Whitney U test, or one of several others. The choice of method depends on the plausibility of normal assumptions, the importance of obtaining a confidence interval, the ease of calculation, etc.

Parametric or non-parametric approaches to evaluation can be applied. Parametric methods are used when the distribution is either known with certainty or can be guessed with a certain degree of certainty. These methods are meaningful only for continuous data which are sampled from a population with an underlying normal distribution or whose distribution can be rendered normal by mathematical transformation. Analysts must be careful to ensure that significant errors are not introduced when assumptions are not met. In contrast, nonparametric methods make no assumptions about the distribution of the data or about the functional form of the regression equation. Nonparametric methods are especially useful in situations where the assumptions required by parametric are in question. Brief overviews and underlying theories of statistical methods that might used in this research are addressed in the following sections.

5.2.1 Parametric Methods

5.2.1.1 The *t*-Test

Student's *t*-test is one of the most commonly used techniques for testing whether the means of two groups are statistically different from each other. This test tries to determine whether the measured difference between two groups is large enough to reject the null hypothesis or whether such differences are just due to "chance". The formula for the *t*-test (Equation 5-1) is a ratio. The numerator of the ratio is just the difference between the two means or averages. The denominator is a measure of the variability or dispersion of the data.

$$\frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{\,} + \frac{s_2^2}{\,}}} \qquad \text{(Equation 5-1)}$$

where $\overline{x_1}$ and $\overline{x_2}$ are the sample means, s_1^2 and s_2^2 are the sample variances, n_1 and n_2 are the sample sizes and t is a Student t quantile with $n_1 + n_2 - 2$ degrees of freedom.

Usually a significance level of 0.05 (or equivalently, 5%) is employed in statistical analyses. The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true. Another index is p-value which is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H_0 is true. The p-value is compared with the actual significance level of the test and, if it is smaller, the result is significant. That is, if the null hypothesis were to be rejected at the 5% significance level, this would be reported as "$p < 0.05$".

The assumptions for *t*-test include: 1) the populations are normally distributed; 2) variances in the two populations are equal; and 3) the populations are independent. The results of the analysis may be incorrect or misleading when assumptions are violated. For example, if the assumption of independence for the sample values is violated, then the two-sample *t* test is simply not appropriate. If the assumption of normality is violated or outliers are present, the two-sample *t* test may not be the most powerful available test. This could mean the difference between detecting a true difference or not. A nonparametric test or employing a transformation may result in a more powerful test.

5.2.1.2 Ordinary Least Squares Regression

Regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that one variable can be predicted from the other, or others (Neter et al. 1996). There are many different kinds of regression models, like the linear regression model, exponential regression model, logistic regression model, and so on. Among them, linear regression is a commonly used and easily understood statistical method. Linear regression explores relationships that can be described by straight lines or their generalization to many dimensions. Regression allows a single response variable to be described by one or more predictor variables.

Ordinary least squares (OLS) regression is a common statistical technique for quantifying the relationship between a continuous dependent variable and one or more independent variables (Neter et al. 1996). The dependent variables may be either continuous or discrete. Neter et al. (1996) provides the basic OLS regression equation for a single variable regression model as shown in Equation 5-2:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_i X_i + \varepsilon_i \quad \text{(Equation 5-2)}$$

where:

$\hat{Y}$	=	value of the response variable in the i^{th} trial
$\hat{\beta}_0$, $\hat{\beta}_i$	=	estimators of regression parameters
X_i	=	value of the predictor variable in the i^{th} trial
ε_i	=	random error term with mean $E\{\varepsilon_i\} = 0$ and variance $\sigma^2 \{\varepsilon_i\}= \sigma^2$; ε_i and ε_j are uncorrelated so that their covariance is zero.

The parameters of the OLS regression equation, β_0 and β_i , are found by the least squares method, which requires that the sum of squares of errors be minimized. Gauss-Markov theorem (Neter et al. 1996) states that, among all unbiased estimators that are linear combinations of *ys*, the OLS estimators of regression coefficients have the smallest variance; i.e., they are the best linear unbiased estimators. The Gauss-Markov Theorem does not tell one to use least squares all the time, but it strongly suggests use of least squares (Neter et al. 1996).

In linear regression, there are key assumptions that must be met, including:

- Y_i are independent normal random variables;
- The expected value of the error terms ε_i is zero;
- The error terms ε_i are assumed to have constant variance $\sigma^{2;}$
- The error terms ε_i are assumed normally distributed;
- The error terms ε_i are assumed to be uncorrelated so that their covariance is zero; and
- The error terms ε_i are independent of the explanatory variable

If the above assumptions are violated the regression equation may yield biased results (Neter et al. 1996). For example, if the explanatory variable is not independent of the error term, larger sample sizes do not lead to lower standard errors for the parameters, and the parameter estimates (slope, etc.) are biased. If the error is not distributed normally, for example, there may be fat tails. Consequently, use of the normal distribution may underestimate true 95% confidence intervals.

5.2.1.3 Robust Regression

OLS models generally rely on the normality assumption and are often fitted by means of the least squares estimators. However, the sensitivity of these estimation techniques is related to this underlying assumption which has been identified as a weakness that can lead to erroneous interpretations (Copt and Heritier 2006). Robust regression procedures dampen the influence of outlying cases, as compared to OLS estimation, in an effort to provide a better fit for the majority of cases. Robust regression procedures are useful when a known, smooth regression function is to be fitted to data that are "noisy", with a number of outlying cases, so that the assumption of a normal distribution for the error terms is not appropriate (Neter et al. 1996). The method of moments (MM) estimators are designed to be both highly robust against outliers and highly efficient.

5.2.2 Nonparametric Methods

Nonparametric methods have several advantages compared with parametric methods. Nonparametric methods require no or very limited assumptions to be made about the format of the data, and they may therefore be preferable when the assumptions required for parametric methods are not valid (Whitley and Ball 2002). Nonparametric methods can be useful for dealing with unexpected, outlying observations that might be problematic with a parametric approach. Nonparametric methods are intuitive and are simple to carry out by hand, for small samples at least.

However, nonparametric methods may lack power as compared with more traditional approaches (Siegel 1988). This lack of power is a particular concern if the sample size is small or if the assumptions for the corresponding parametric method hold true (e.g., normality of the data). Nonparametric methods are geared toward hypothesis testing rather than estimation of effects. It is often possible to obtain nonparametric estimates and associated confidence intervals, but this process is not generally straightforward. In addition, appropriate computer software for nonparametric methods can be limited, although the situation is improving.

5.2.2.1 Chi-Square Test

The Chi-square (Koehler and Larnz 1980), best known goodness-of-fit test, assumes that the observations are independent and that the sample size is reasonably large. This method can be used to test whether a sample fits a known distribution, or whether two unknown distributions from different samples are the same. The test can detect major departures from a logistic response function, but is not sensitive to small departures from a logistic response function. The test assumptions are that the sample is random and that the measurement scale is at least ordinal (Conover 1980; Neter et al. 1996).

Pearson's chi-square goodness of fit test statistic is shown in Equation 5-3 (StatsDirect 2005):

$$T = \sum_{j=1}^{i} \frac{\left(O_j - E_j\right)^2}{E_j} \qquad \text{(Equation 5-3)}$$

where O_j are observed counts, E_j are corresponding expected count and c is the number of classes for which counts/frequencies are being analyzed.

The test statistic is distributed approximately as a chi-square random variable with c-1 degrees of freedom. The test has relatively low power (chance of detecting a real effect) with all but large numbers or big deviations from the null hypothesis (all classes contain observations that could have been in those classes by chance).

The handling of small expected frequencies is controversial. Koehler and Larnz asserted that the chi-square approximation is adequate provided all of the following are true: total of observed counts (N) ≥ 10; number of classes (c) ≥ 3; all expected values ≥ 0.25 (Koehler and Larnz 1980).

5.2.2.2 Kolmogorv-Smirnov Two-Sample Test

The Kolmogorov-Smirnov (K/S) two-sample test (Chakravart and Roy 1967) compares the empirical distribution functions of two samples, E_1 and E_2. The Kolmogorov-Smirnov test is a nonparametric test, which can be used to test whether two or more samples are governed by the same distribution by comparing their empirical distribution functions.

The Kolmogorov-Smirnov two sample test statistic can be defined as shown in Equation 5-4 (Chakravart and Roy 1967):

$$D = \left| E_1(i) - E_2(i) \right| \qquad \text{(Equation 5-4)}$$

where E_1 and E_2 are the empirical distribution functions for the two samples.

The Kolmogorov-Smirnov (K/S) two-sample test provides an improved methodology over the chi-squared test since data do not have to be assigned arbitrarily to bins. Further, it is a non-parametric test so a distribution does not have to be assumed. However, the main disadvantage to the K/S is similar to the chi-square in that the orders of magnitude of separate tests that would have to be conducted to test all the possible combinations of variables in the datasets is logistically infeasible (Hallmark 1999).

5.2.2.3 Wilcoxon Mann-Whitney Test

The Wilcoxon Mann-Whitney Test (Easton and McColl 2005) is one of the most powerful of the nonparametric tests for comparing two populations. This test is used to test the null hypothesis that two populations have identical distribution functions against the alternative hypothesis that the two distribution functions differ only with respect to location (median), if at all.

The Wilcoxon Mann-Whitney test does not require the assumption that the differences between the two samples are normally distributed. In many applications, the Wilcoxon Mann-Whitney Test is used in place of the two sample *t*-test when the normality assumption is questionable. This test can also be applied when the observations in a sample of data are ranks, that is, ordinal data rather than direct measurements.

The Mann Whitney U statistic is defined as shown in Equation 5-5 (StatsDirect 2005):

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i-n_1+1}^{n_2} R_i \quad \text{(Equation 5-5)}$$

where samples of size n_1 and n_2 are pooled and R_i are the ranks.

U can be resolved as the number of times observations in one sample precede observations in the other sample in the ranking. Wilcoxon rank sum, Kendall's S and the Mann-Whitney U test are exactly equivalent tests. In the presence of ties the Mann-Whitney test is also equivalent to a chi-square test for trend.

5.2.2.4 Analysis of Variance (ANOVA)

ANOVA (Analysis of Variance) (Neter et al. 1996), sometimes called an F test, is closely related to the *t* test. The major difference is that, where the *t* test measures the difference between the means of two groups, an ANOVA tests the difference between the means of two or more groups. ANOVA modeling does not require any assumptions about the nature of the statistical relation between the response and explanatory variables, nor do they require that the explanatory variables be quantitative.

The ANOVA, or single factor ANOVA, compares several groups of observations, all of which are independent, but each group of observations may have a different mean. A test of great importance is whether or not all the means are equal. The advantage of using ANOVA rather than multiple *t*-tests is that it reduces the probability of a type-I error (making multiple comparisons increases the likelihood of finding something by chance). One potential drawback to an ANOVA is that it can only tell that there is a significant difference between groups, not which groups are significantly different from each other. The breakdowns of the total sum of squares and degrees of freedom, together with the resulting mean squares, are presented in an ANOVA table such as Table 5-1.

Table 5-1 ANOVA Table for Single-Factor Study (Neter et al. 1996)

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	Expected Mean Square E(MS)
Between treatments	$SSTR = \sum n_i \left(\overline{Y}_{i} - \overline{Y}_{..}\right)^2$	r - 1	$MSTR = \dfrac{SSTR}{r-1}$	$\sigma^2 + \dfrac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$
Error (within treatments)	$SSE = \sum\sum \left(Y_{ij} - \overline{Y}_{i}\right)^2$	n_T - r	$MSE = \dfrac{SSE}{N_T - r}$	σ^2
Total	$SSTO = \sum\sum \left(Y_{ij} - \overline{Y}_{..}\right)^2$	n_T - 1		

A factorial ANOVA can examine data that are classified on multiple independent variables. A factorial ANOVA can show whether there are significant main effects of the independent variables and whether there are significant interaction effects between independent variables in a set of data. Interaction effects occur when the impact of one independent variable depends on the level of the second independent variable (Neter et al. 1996). Computation can be performed with standard statistical software such as SAS®.

5.2.2.5 HTBR

HTBR (Breiman et al. 1984) is a forward step-wise variable selection method, similar to forward stepwise regression. This method is also known as Classification and Regression Tree (CART) analysis. This technique generates a "tree" structure by dividing the sample data recursively into a number of groups. The groups are selected to maximize some measure of difference in the response variable in the resulting groups. As Washington et al. summarized in 1997 (Washington et al. 1997a), this method is based upon iteratively asking and answering the following questions: (1) which variable of all of the variables 'offered' in the model should be selected to produce the maximum reduction in variability of the response? and (2) which value of the selected variable (discrete or continuous) results in the maximum reduction in variability of the response? The HTBR terminology is similar to that of a tree; there are branches, branch splits or internal nodes, and leaves or terminal nodes (Washington et al. 1997a).

To explain the method in mathematical terms, the definitions are presented by Washington et al. (Washington et al. 1997a). The first step is to define the deviance at a node. A node represents a data set containing L observations. The deviance, D_a , can be estimated as shown in equation 5-6:

$$D_a = \sum_{l=1}^{L} (y_{l,a} - \overline{x}_a)^2 \qquad \text{(Equation 5-6)}$$

where

D_a	=	total deviance at node a, or the sum of squared error (SSE) at the node
$y_{l,a}$	=	l^{th} observation of dependent variable y at node *a*
$\overline{x}_a$	=	estimated mean of L observations in node *a*

Next, the algorithm seeks to split the observation at node a on a value of an independent variable, X_i , into two branches and corresponding nodes *b* and *c*, each containing M and N of the original L observations (M+N=L) of the variable X_i. The deviance reduction function evaluated over all possible *Xs* then can be defined as shown in Equations 5-7 thru 5-9:

$$\Delta_{(allX)} = D_a - D_b - D_c \qquad \text{(Equation 5-7)}$$

$$D_b = \sum_{m=1}^{M} (y_{m,b} - \overline{x}_b)^2 \qquad \text{(Equation 5-8)}$$

$$D_c = \sum_{n=1}^{N} (y_{n,c} - \overline{x}_c)^2 \qquad \text{(Equation 5-9)}$$

where

$\Delta_{(allX)}$ = the total deviance reduction function evaluated over the domain of all Xs

D_b = total deviance at node b

D_c = total deviance at node c

$y_{m,b}$ = m^{th} observation on dependent variable y in node b

$y_{n,c}$ = n^{th} observation on dependent variable y in node c

$\overline{x}_b$ = estimated mean of M observations in node b

$\overline{x}_c$ = estimated mean of N observations in node c

The variable X_k and its optimum split $X_{k(i)}$ is sought so that the reduction in deviance is maximized, or more formally when (as shown in equation 5-10):

$$\Delta_{(allX)} = \sum_{l=1}^{L}(y_{l,a} - \overline{x}_a)^2 - \sum_{m=1}^{M}(y_{m,b} - \overline{x}_b)^2 - \sum_{n=1}^{N}(y_{n,c} - \overline{x}_c)^2 = \max \qquad \text{(Equation 5-10)}$$

where

$\Delta_{(allX)}$ = the total deviance reduction function evaluated over the domain of all Xs

$y_{l,a}$ = l^{th} observation of dependent variable y at node *a*

$\overline{x}_a$ = estimated mean of L observations in node *a*

$y_{m,b}$ = m^{th} observation on dependent variable y in node b

$y_{n,c}$ = n^{th} observation on dependent variable y in node c

$\overline{x}_b$ = estimated mean of M observations in node b

$\overline{x}_c$ = estimated mean of N observations in node c

The maximum reduction occurs at a specific value $X_{k(i)}$, of the independent variable X_k. When the data are split at this point, the remaining samples have a much smaller variance than the original data set. Thus, the reduction in node *a* deviance is greatest when the deviances at nodes *b* and *c* are smallest. Numerical search procedures are employed to maximize Equation 5-10 by varying the selection of variables used as a basis for a split and the value to use for each variable at a split.

In growing a regression tree, the binary partitioning algorithm recursively splits the data in each node until the node is homogenous or the node contains too few observations. If left unconstrained, a regression tree model can "grow" until it results in a complex model with a single observation at each terminal node that explains all the deviance. However, for application purposes, it is desirable to create criteria to balance the model's ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. Some software, such as S-Plus™, allows the user to select such criteria. The software allows the user to interact with the data in the following manner to select variables and help simplify the final model:

- Response variable: the response variable is selected by the user from a list of fields from the data set;
- Predictor variables: one or more independent variables can be selected by the user from a list of fields associated with the dataset;
- Minimum number of observations allowed at a single split: sets the minimum number of observations that must be present before a split is allowed (default is 5);
- Minimum node size: sets the allowed sample size at each node (default is 10); and
- Minimum node deviance: the deviance allowed at each node (default is 0.01).

However, unlike OLS regression models, a shortcoming of HTBR is the absence of formal measures of model fit, such as *t*-statistics, F-ratio, and r-square, to name a few. Thus, the HTBR model is used to guide the development of an OLS regression model, rather than as a model in its own right. Similar uses of HTBR techniques have been developed and applied in previous research papers (Washington et al. 1997a; Washington et al. 1997b; Fomunung et al. 1999; Frey et al. 2002).

5.3 Modeling Approach

The model development process will start by using HTBR both as a data reduction tool and for identifying potential interactions among the variables. Then OLS Regression or Robust Regression is used with the identified variables to estimate a preliminary "final" model. After that, we need to check the model for compliance with normality assumptions and goodness of fit.

Several diagnostic tools are available to perform these checks. Once a preliminary "final" model is obtained, regression coefficients are examined using their *t*-statistics and correlation coefficients to determine which variables should be removed or retained in the model for further analysis. However this procedure can lead to the removal of potentially important inter-correlated explanatory variables. In fact, variable agreement with underlying scientific principles of combustion, pollutant formation and emission controls (cause-effect relationships) should be the basis for the ultimate decisions regarding variable selection. Thus, a *t*-statistic may indicate that a parameter is insignificant (at level of significance = 0.05), while theory indicates that such a parameter should be retained in the model for further analysis. This type of error is usually referred to as a type II error (Fomunung 2000).

F-statistics and adjusted coefficient of multiple determination, R^2 are used to determine the effect-size of the parameters. Usually, adding more explanatory variables to the regression model can only increase R^2 and never reduce it, because SSE can never become larger with more X variables and total sum of squares (SSTO) is always the same for a given set of responses. The adjusted coefficient of multiple determination can adjust R^2 by dividing each sum of squares by its associated degrees of freedom. The F-test is used to test whether the parameter can be dropped even if the *t*-statistic is appropriate.

In multiple regression analysis, the predictor or explanatory variables tend to be correlated among themselves and with other variables related to the response variable but not included in the model. The effects of multicollinearity are many and can be severe. Neter et al. (Neter et al. 1996) have documented a few of these: when multicollinearity exists the interpretation of partial slope coefficients becomes meaningless; multicollinearity can lead to estimated regression coefficients that vary widely from one sample to another; and there may be several regression functions that provide equally good fits to the data, making the effects of individual predictor variables difficult to assess.

There are some informal diagnostic tools suggested to detect this problem. A frequently used technique is to calculate a simple correlation coefficient between the predictor variables to detect the presence of inter-correlation among independent variables. Large values of correlation is an indication that multicollinearity may exist. Large changes in the estimated regression coefficients when a predictor variable is added or deleted are also an indication. Finally, multicollinearity may be a problem if estimated regression coefficients are calculated with an algebraic sign that is the opposite of that expected from theoretical considerations or prior experience (i.e., the beta coefficient is compensating for the beta coefficient of a correlated explanatory variable).

A formal method of detecting this problem is the variance inflation factor (VIF), which is a measure of how much the variances of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related (Neter et al. 1996). This method is widely used because it can provide quantitative measurements of the impact of multicollinearity. The largest VIF value among all Xs is used to assess the severity of multicollinearity. As a rule of thumb, a VIF in excess of 10 is frequently used as an indication that multicollinearity is severe.

Diagnostic plots are examined to verify normality and homoscedasticity (i.e., homogeneity of variance) assumptions as well as the goodness of fit. Because of the difficulty in assessing normality, it is usually recommended that non-constancy of error variance should be investigated first (Neter et al. 1996). The plots used to identify any patterns in the residuals are considered as informal diagnostic tools and include plots of the residuals versus the fitted values and plot of square root of absolute residuals versus the fitted values. The normality of the residuals can be studied from histograms, box plots, and normal probability plots of the residuals. In addition, comparisons can be made of observed frequencies with expected frequencies if normality exists. Usually, heteroscedasticity and/or inappropriate regression functions may induce a departure from normality. When OLS is applied to heteroskedastic models the estimated variance is a biased estimator of the true variance. OLS either overestimates or underestimates the true variance, and, in general it is not possible to determine the nature of the bias. The variances, and the standard errors, may therefore be either understated or overstated.

5.4 Model Validation

Model validity refers to the stability and reasonableness of the regression coefficients, the plausibility and usability of the regression function, and the ability to generalize inferences drawn from the regression function. Validation is a useful and necessary part of the model-building process (Neter et al. 1996).

Two basic ways of validating a regression model are internal and external. Internal validation consists of model checking for plausibility of signs and magnitudes of estimated coefficients, agreement with earlier empirical results and theory, and model diagnostic checks such as distribution of error terms, normality of error terms, etc. Internal validation will be performed as part of the model estimation procedure.

External validation is the process to check the model and its predictive ability with the collection of new data, such as data from another location or time, or using a holdout sample. Considering there are only 15 buses/engines in the data set, it is not practical to split the data set and hold a sample for validation purposes. Splitting the data set will definitely influence the regression estimators. However suggestions and procedures for external validation will be provided.

CHAPTER 6

6. DATA SET SELECTION AND ANALYSIS OF EXPLANATORY VARIABLES

6.1 Data Set Used for Model Development

Development of a modal model designed to predict emissions on a second-by-second basis as a function of engine load requires the availability of appropriate emission test data. Modal modeling required the availability of second-by-second vehicle emissions data, collected in parallel with corresponding revealed engine load data. In 2004, only two data sets could be identified for use in this modeling effort. U.S. EPA provided two major HDV activity and emission databases to develop the emission rate model (Ensfield 2002) (U.S. EPA 2001b). One database is a transit bus database, which included emissions data collected on diesel transit buses operated by the AATA in 2001, and another database is heavy HDV (HDV8B) database prepared by NRMRL in 2001. The transit database consisted of data collected from 15 buses with the same type of engines while the HDV8B database consisted of only one truck engine tested extensively on-road under pre-rebuild and post rebuild engine conditions. To decide whether it is suitable to combine these two data sets or treat them individually, two dummy variables were added to the databases to describe vehicle types. For the first dummy variable named "bus", 1 was assigned for transit bus, and 0 for others. For the second dummy variable, 1 was assigned for HDDV with pre-rebuild engine, and 0 for others. HTBR was applied to all data sets to examine whether transit buses behave differently from HDDVs or not. The regression trees and results for NO_x, CO, and HC emission rates are given in Figures 6-1 to Figure 6-3.

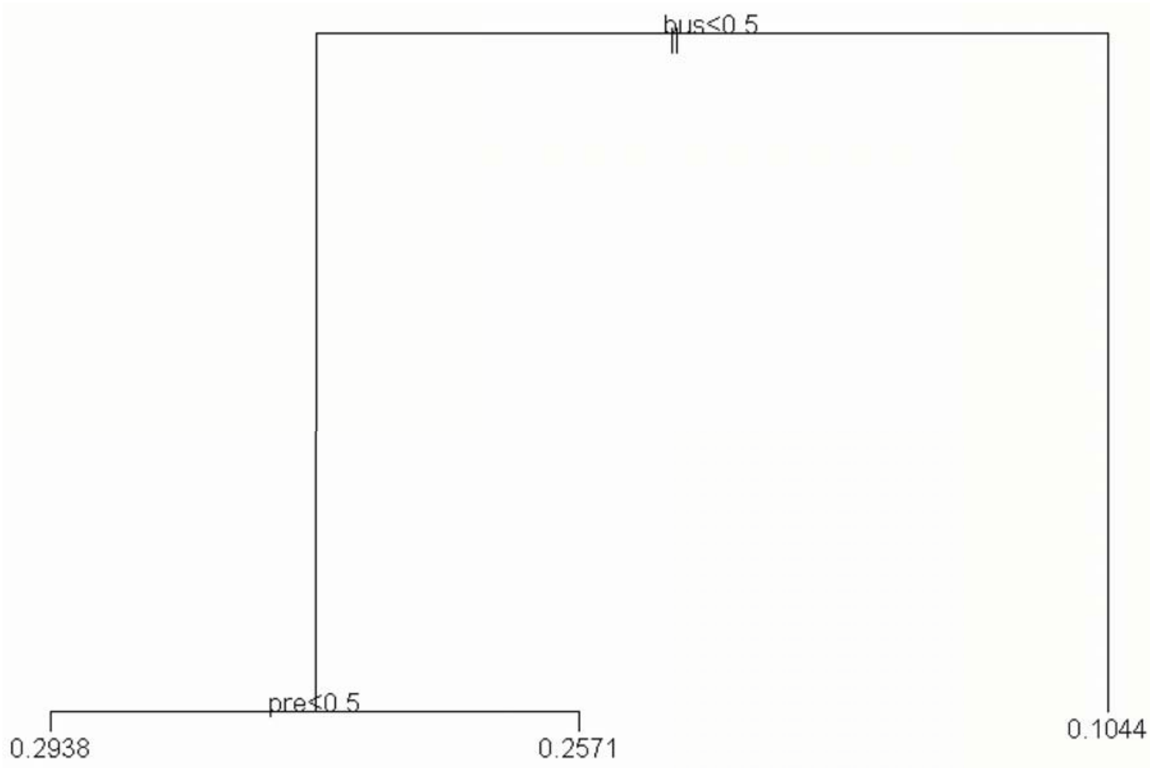


Figure 6-1 HTBR Regression Tree Result for NO_x Emission Rate for All Data Sets

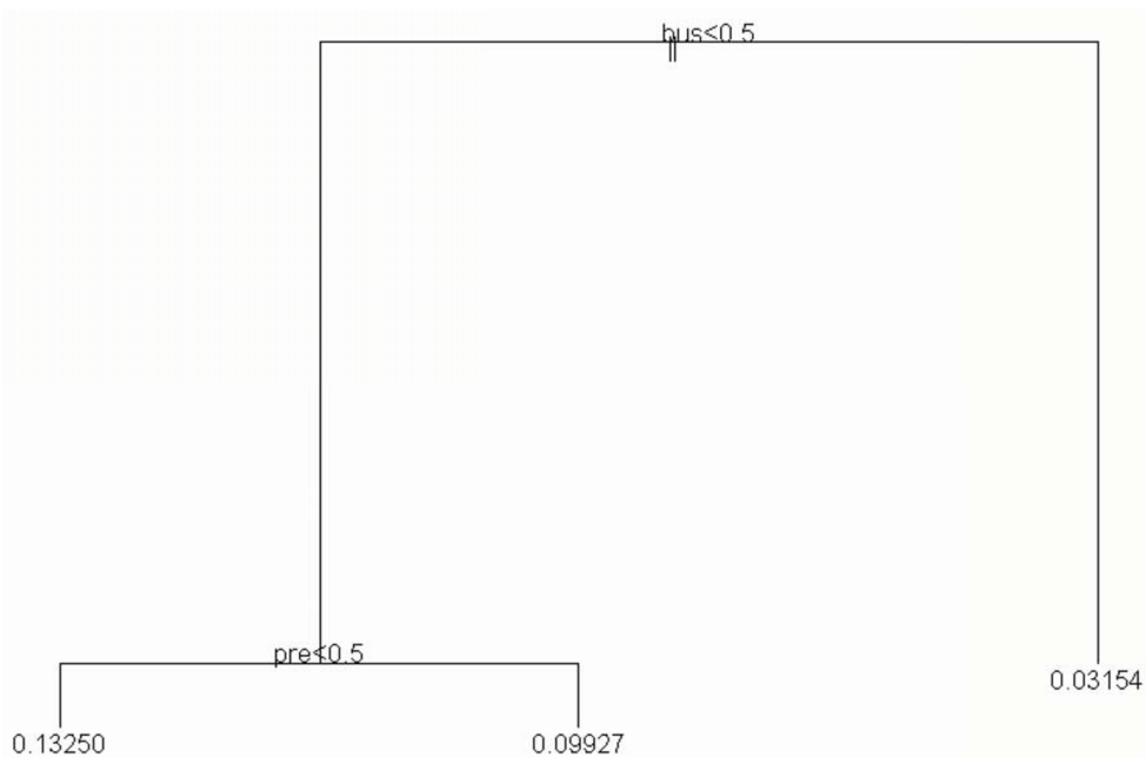


Figure 6-2 HTBR Regression Tree Result for CO Emission Rate for All Data Sets

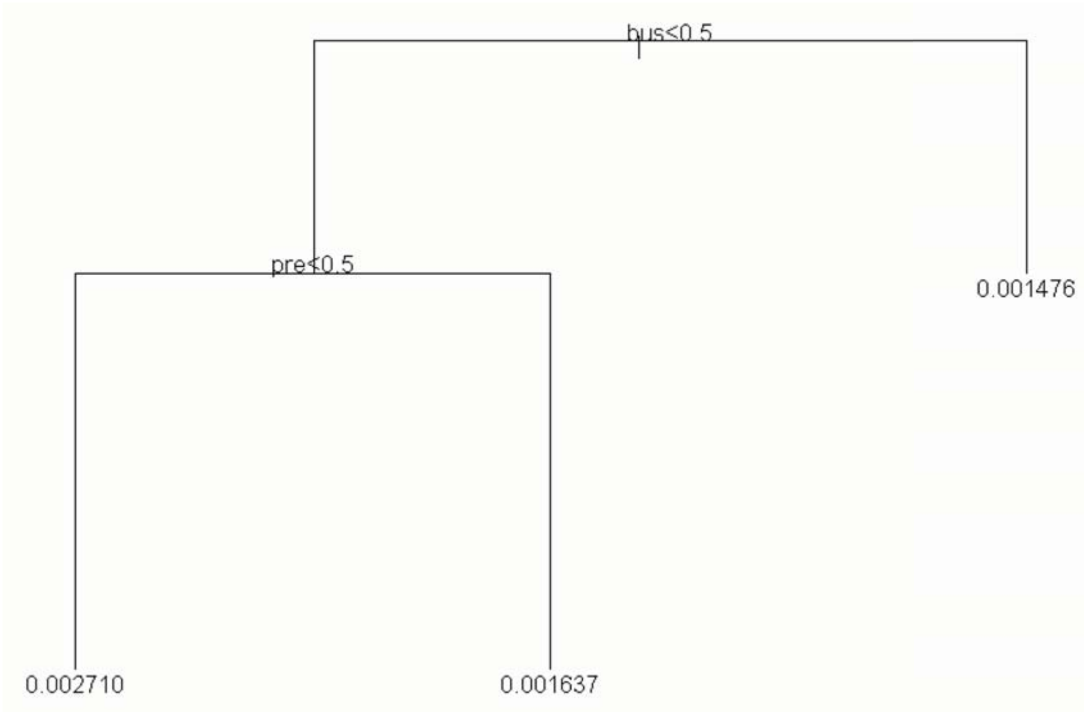


Figure 6-3 HTBR Regression Tree Result for HC Emission Rate for All Data Sets

Dummy variable for bus is selected as the first split for all three trees above. Therefore transit bus and HDDV should be treated separately. Since there are 15 engines in the transit bus data set and one engine (pre-rebuild and post-rebuild for the same engine) in the HDDV data set, the transit bus data set should be used for the final version of the conceptual model development.

6.2 Representative Ability of the Transit Bus Data Set

The transit bus data set was collected by Sensors, Inc. in Oct. 2001 (Ensfield 2002). The buses tested came from the AATA and included 15 New Flyer models with Detroit Diesel Series 50 engines. All of the buses were of model years 1995 and 1996. All of the bus tested periods lasted approximately 2 hours. The buses operated during standard AATA bus routes and stopped at all regular stops although the buses did not board or discharge any passengers (Ensfield 2002) The routes were mostly different for each test, and were selected for a wide variety of driving conditions (see Figure 4-1).

Figure 6-4 shows the speed-acceleration matrix developed with second-by-second data. There are two high speed/acceleration frequency peaks here. One is the bin of speed ≤ 2.5 mph and acceleration (-0.25 mph/s, 0.25 mph/s) and contains 26.11% of the observations, while the other is the combination of several adjacent bins which covers speed (22.5 mph, 47.5 mph) and acceleration (-0.75 mph/s, 0.75 mph/s).

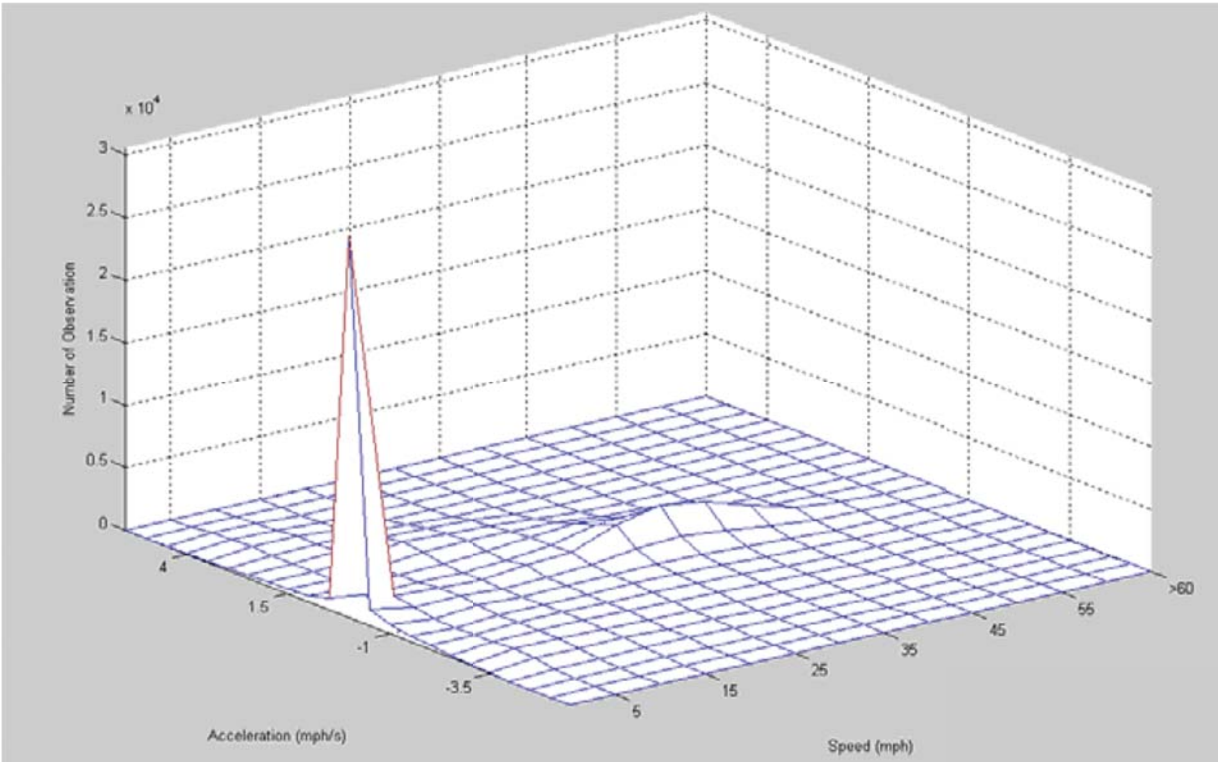


Figure 6-4 Transit Bus Speed-Acceleration Matrix

Georgia Institute of Technology researchers collected more than 6.5 million seconds of transit bus speed and position data using Georgia Tech Trip Data Collectors (an onboard computer with GPS receiver, data storage, and wireless communication device) installed on two Metropolitan Atlanta Rapid Transit Authority (MARTA) buses in 2004 (Yoon et al. 2005b). With second-by-second data, the research team developed transit bus speed/acceleration matrices for the combinations between roadway facility type (arterial or local road) and time range (morning, midday, afternoon, night). For each matrix, two high acceleration/deceleration frequency peaks were also found. This finding is consistent with the AATA data set, indicating at least that the on-road operations of the buses in Ann Arbor are similar to operations in the Atlanta region.

This data set was collected under a wide variety of environmental conditions, too. The temperature ranged from 10 °C to 30 °C, the relative humidity ranged from 15% to 65%, while the barometric pressure ranged from 960 mbar to 1000 mbar (Figure 6-5). So we can use this data set to examine the impact of environmental conditions on emissions.

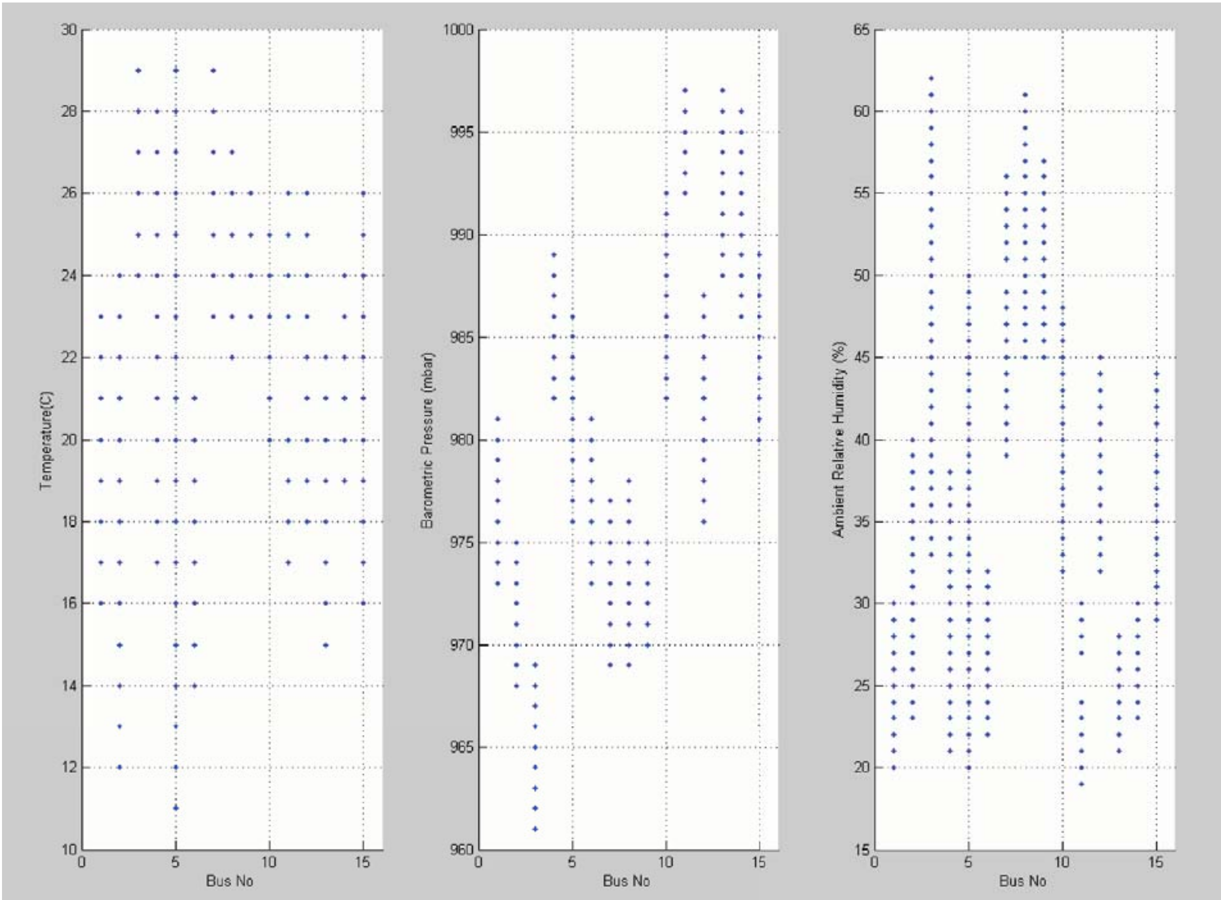


Figure 6-5 Test Environmental Conditions

Transit buses tested were provided by the AATA and all of them are New Flyer models with Detroit Diesel Series 50 engines. Since these buses utilized consistent engine technologies (i.e., fuel injection type, catalytic converter type, transmission type, and so on), the ability of estimated emission models to incorporate the effect of other types of vehicle technologies is limited. Another limitation is the consideration of the effects of emission control technology deterioration on emission levels since these buses were only 5 or 6 years old during the test.

6.3 Variability in Emissions Data

6.3.1 Inter-bus Variability

Data are presented to illustrate the variability in observed data. Inter-bus variabilities are illustrated using median and mean of NO_x, CO, and HC emission rates for each bus from Figures 6-6 to 6-8. The difference between median and mean is an indicator of skewness for the distribution of emission rates.

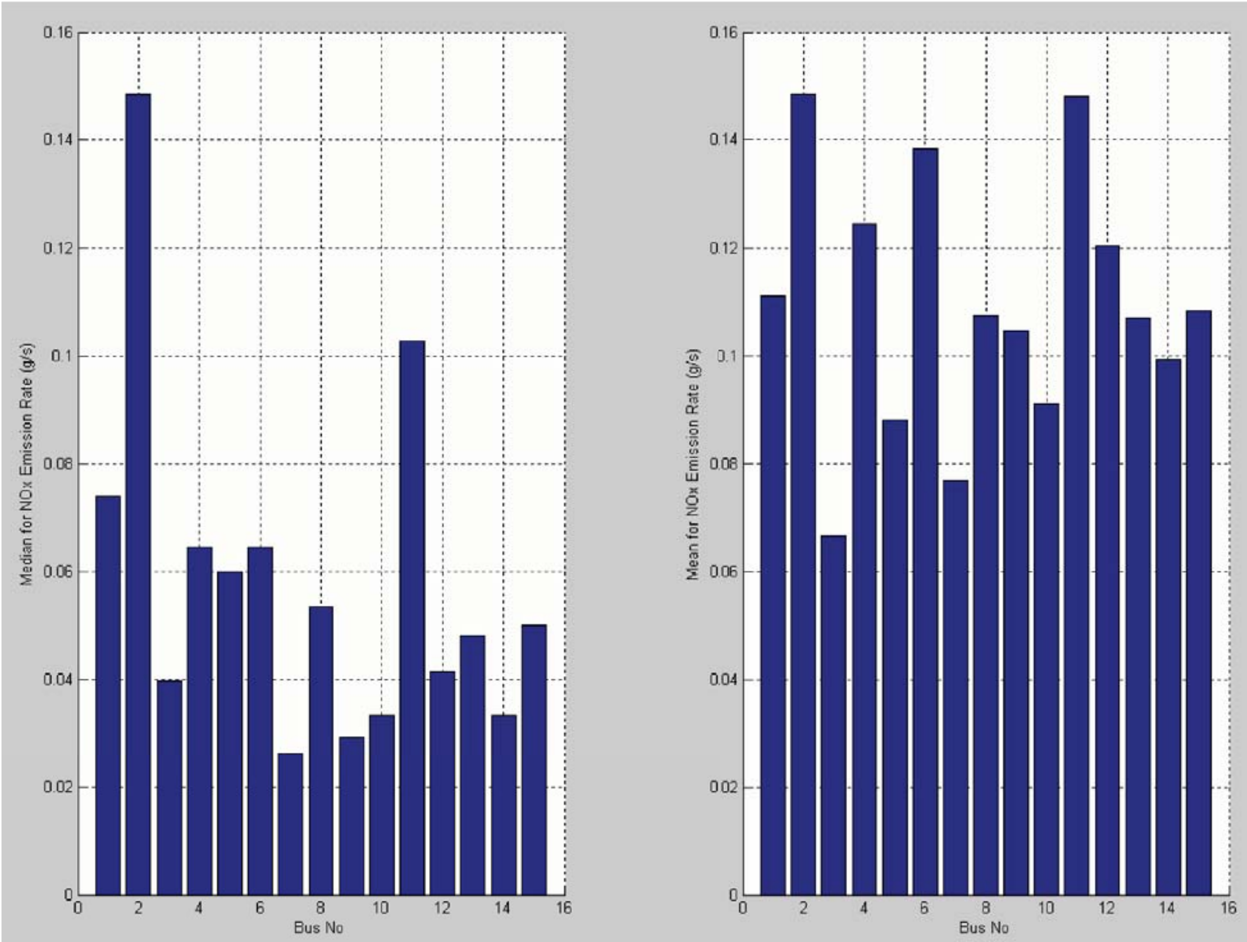


Figure 6-6 Median and Mean of NO_x Emission Rates by Bus

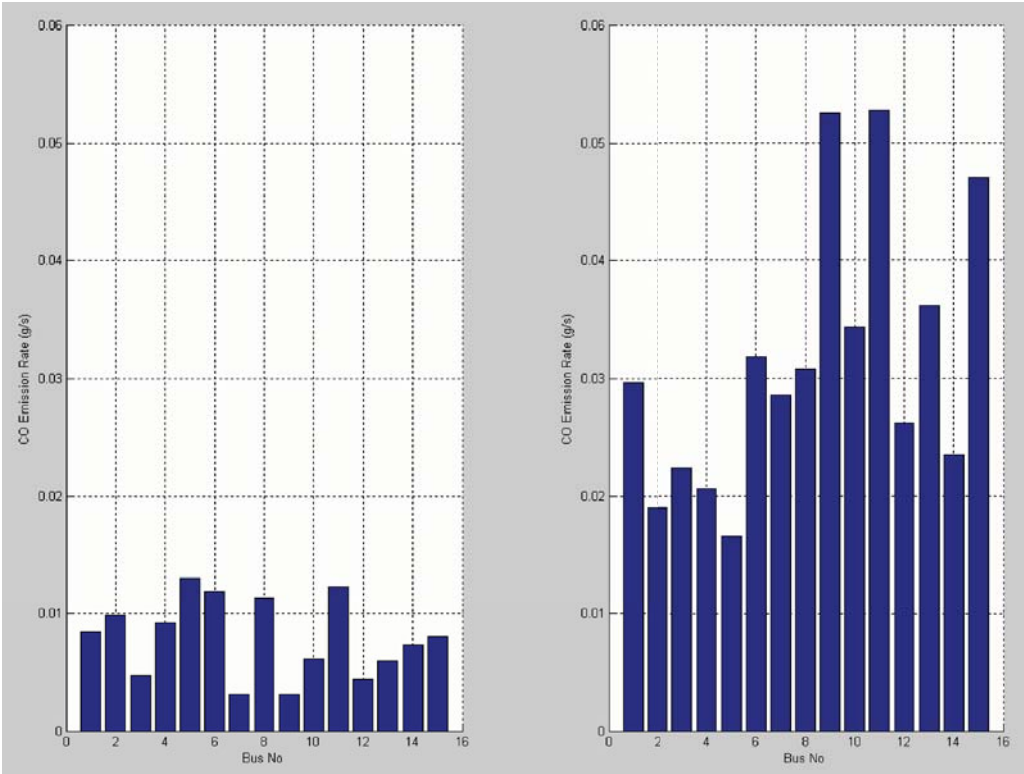


Figure 6-7 Median and Mean of CO Emission Rates by Bus

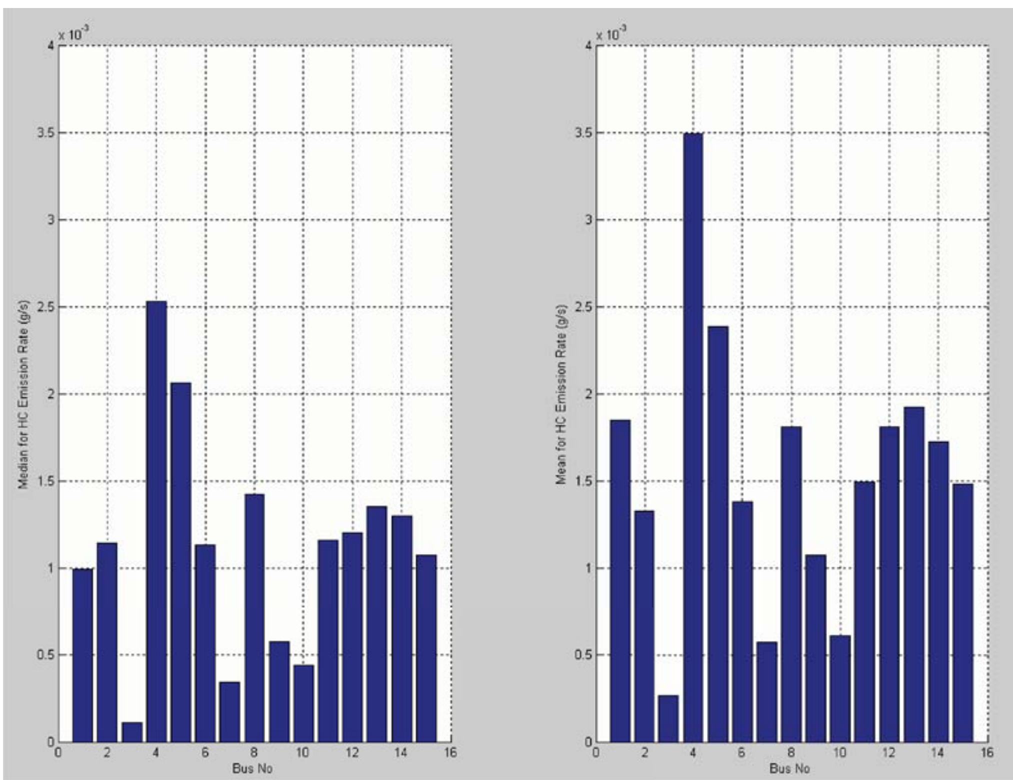


Figure 6-8 Median and Mean of HC Emission Rates by Bus

The purpose of inter-bus variability analysis was to characterize the range of variability in vehicle average emissions among all of the buses, to determine whether the data set is relatively homogeneous. Although there are some clusters among the buses as suggested from Figures 6-6 to 6-8 and some skewness in the distribution as suggested by upper tails in Figure 6-9, it is not obvious that this data set lacks homogeneity and should be separated into different groups. Thus, this data set is treated as a single group for purposes of analysis and model development.

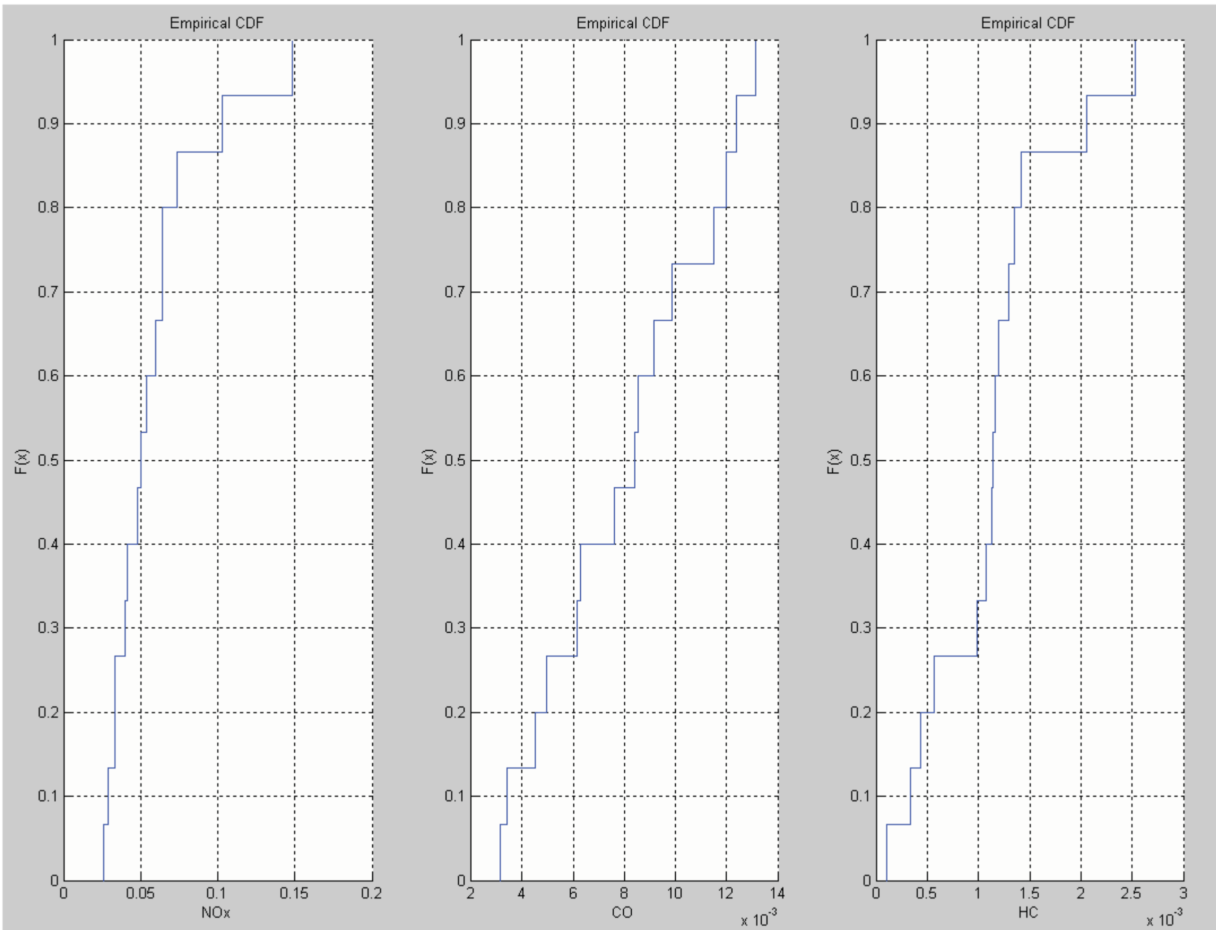


Figure 6-9 Empirical Cumulative Distribution Function Based on Bus Based Median Emission Rates for Transit Buses

6.3.2 Descriptive Statistics for Emissions Data

Applicable numerical summary statistics, such as variable means and standard deviations, are presented in Table 6-1. Relatively simple graphics such as histograms and boxplots describing variable distributions are presented in Figures 6-10 to 6-12. It may also be necessary to assess whether the individual variables are normally distributed prior to any further analysis using parametric methods that are based upon this assumption.

Table 6-1 Basic Summary Statistics for Emissions Rate Data for Transit Bus

*** Summary Statistics for data in: transitbus.data ***			
	NO_x	HC	CO
Min:	0.000000e+000	0.000000e+000	0.000000e+000
1st Qu.:	3.030000e-003	2.195000e-002	4.200000e-004
Mean:	3.183675e-002	1.052101e-001	1.438709e-003
Median:	7.540000e-003	5.058000e-002	9.300000e-004
3rd Qu.:	2.197000e-002	1.731100e-001	1.840000e-003
Max:	3.057700e+000	2.427900e+000	6.679000e-002
Total N:	1.075350e+005	1.075350e+005	1.075350e+005
NA's :	0.000000e+000	0.000000e+000	0.000000e+000
Std Dev.:	8.479305e-002	1.162344e-001	1.956353e-003

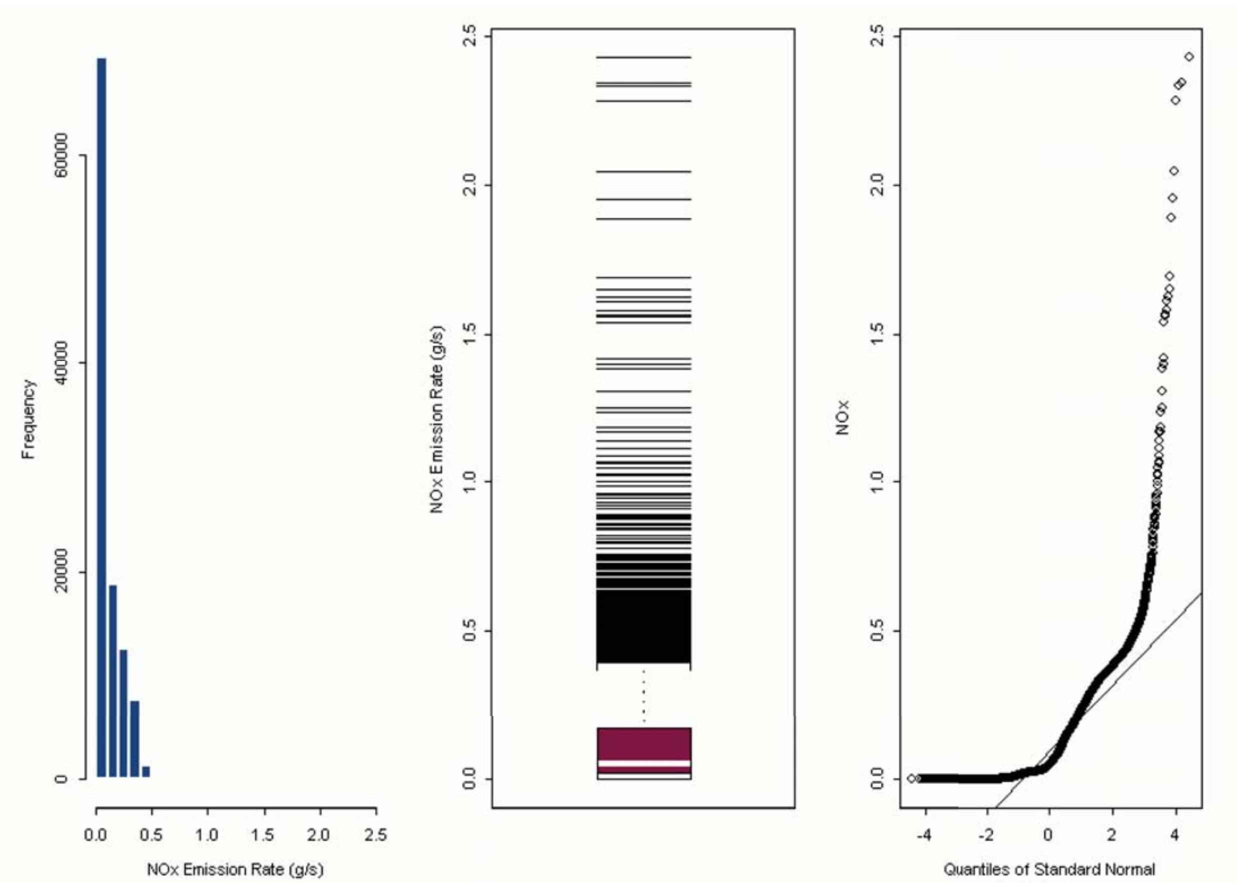


Figure 6-10 Histogram, Boxplot, and Probability Plot of NO_x Emission Rate

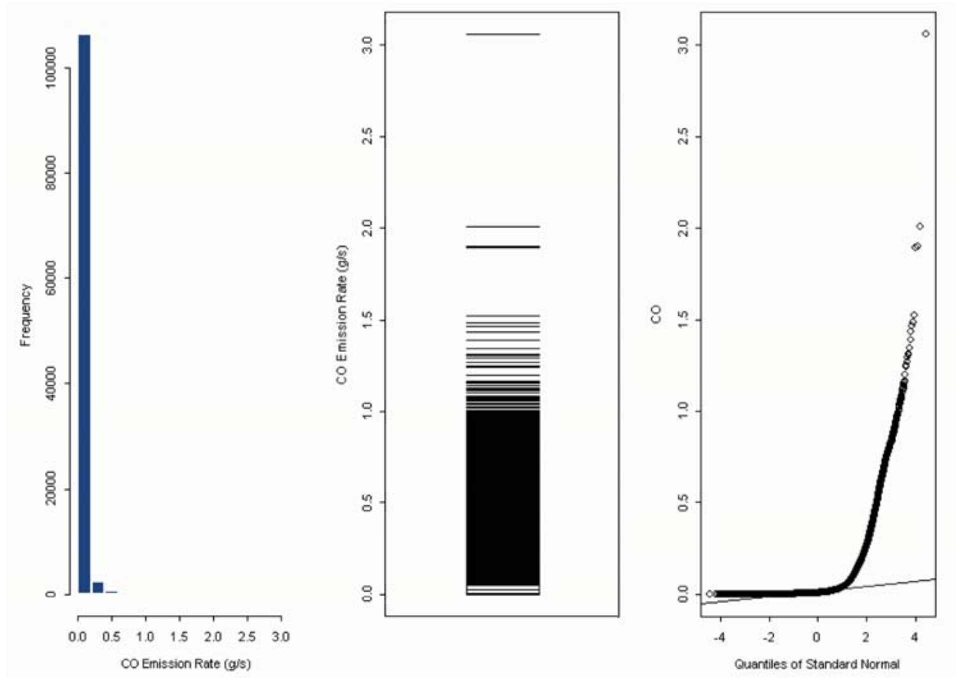


Figure 6-11 Histogram, Boxplot, and Probability Plot of CO Emission Rate

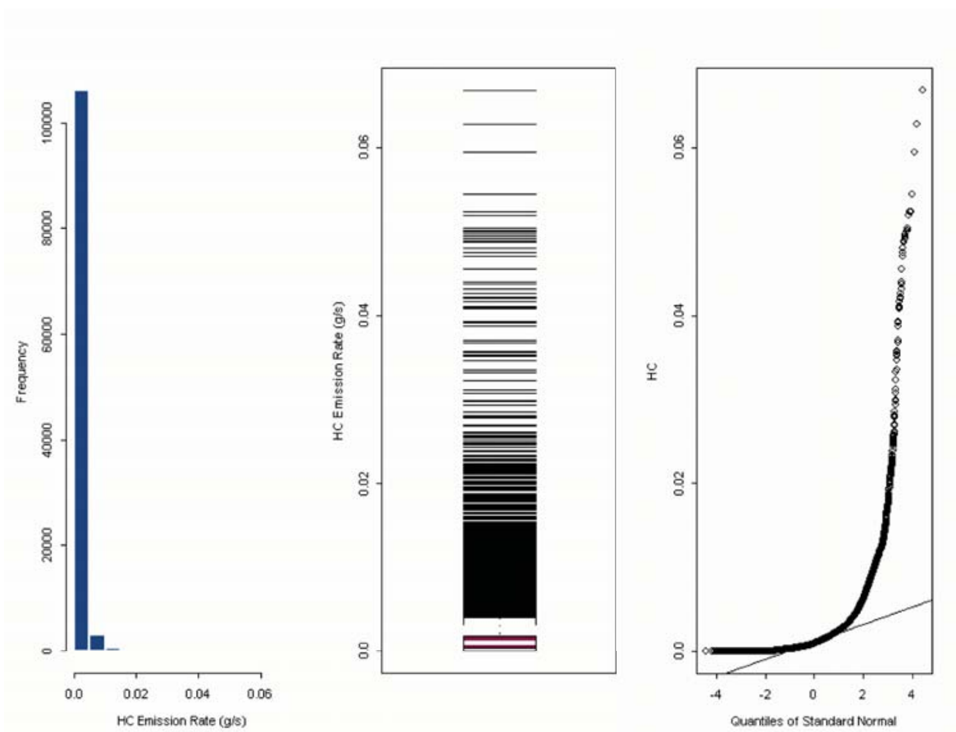


Figure 6-12 Histogram, Boxplot, and Probability Plot of HC Emission Rate

Further analysis indicated that there are some zero values in the emission data. There might be several reasons for zero values. Missing data caused by loss of communication between instruments or failure of a particular vehicle were recorded as zero in the data set. Those zero values were already identified in the data post-processing procedure in Chapter 4. Zero values might also have occurred when the reference air contained significant amounts of a pollutant so the instrument systematically reported negative emission values. Sensors, Inc. suggested that negative data should be set to zero. Thus these negative values were artificially recorded as zero, not observed by test equipment as zero. These zero values would create truncation issues in the model, since the Sensors, Inc. transit bus data set contained only valid positive emission data. Usually, truncation is found when a random variable is not observable over its entire range. Truncation could not be treated as a missing data problem as the missing observations are random. In statistics consideration or analysis can be limited to data that meet certain criteria or to a data distribution where values above or below a certain point have been eliminated (or cannot occur). A program was written in MATLAB® to check for the presence of zero emissions estimates in the data set. There were 1.45% zero values for NO_x emissions, 1.65% zero values for CO emissions and 3.84% zero values for HC emissions. Since negative emission values were not observable for the transit bus data set, further analysis will focus on truncated data sets with valid positive emission data only.

The numerical summary statistics such as variable means and standard deviations for truncated emission data are presented in Table 6-2, and relatively simple graphics such as histograms and boxplots describing variable distributions are presented from Figures 6-13 to 6-15. The mean of truncated NO_x emission data increases 1.26%, while the mean of truncated CO emission data increases 1.23% and the mean of truncated HC emission data increases 0.99%, compared with the means of the original data set.

Table 6-2 Basic Summary Statistics for Truncated Emissions Rate Data

	NO_x	CO	HC
Min:	1.000000e-005	1.000000e-005	1.000000e-005
1st Qu.:	2.256000e-002	3.190000e-003	4.700000e-004
Mean:	1.067578e-001	3.236955e-002	1.496171e-003
Median:	5.243500e-002	7.770000e-003	9.900000e-004
3rd Qu.:	1.749625e-001	2.246000e-002	1.880000e-003
Max:	2.427900e+000	3.057700e+000	6.679000e-002
Total N:	1.059760e+005	1.057650e+005	1.034050e+005
NA's :	0.000000e+000	0.000000e+000	0.000000e+000
Std Dev.:	1.163785e-001	8.539871e-002	1.973375e-003

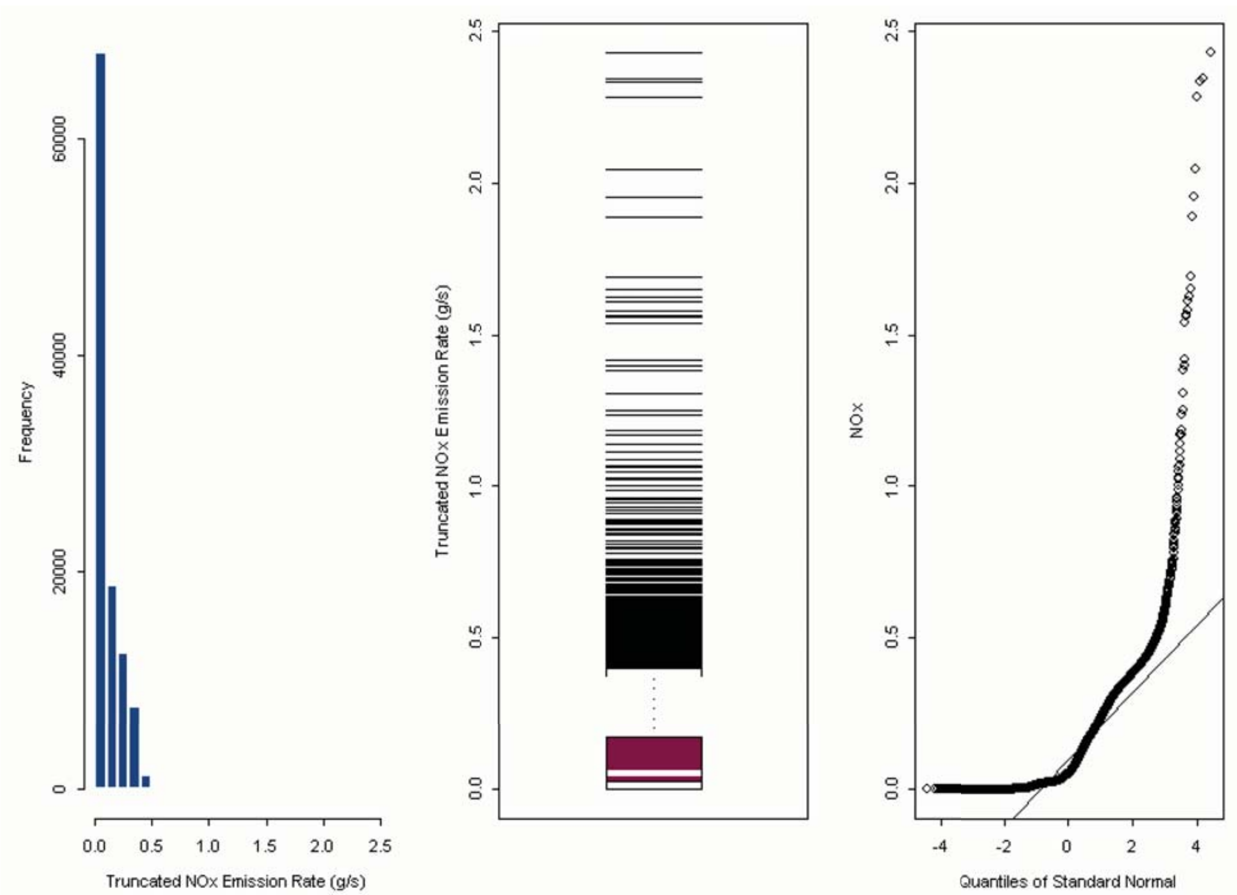


Figure 6-13 Histogram, Boxplot, and Probability Plot of Truncated NO_x Emission Rate

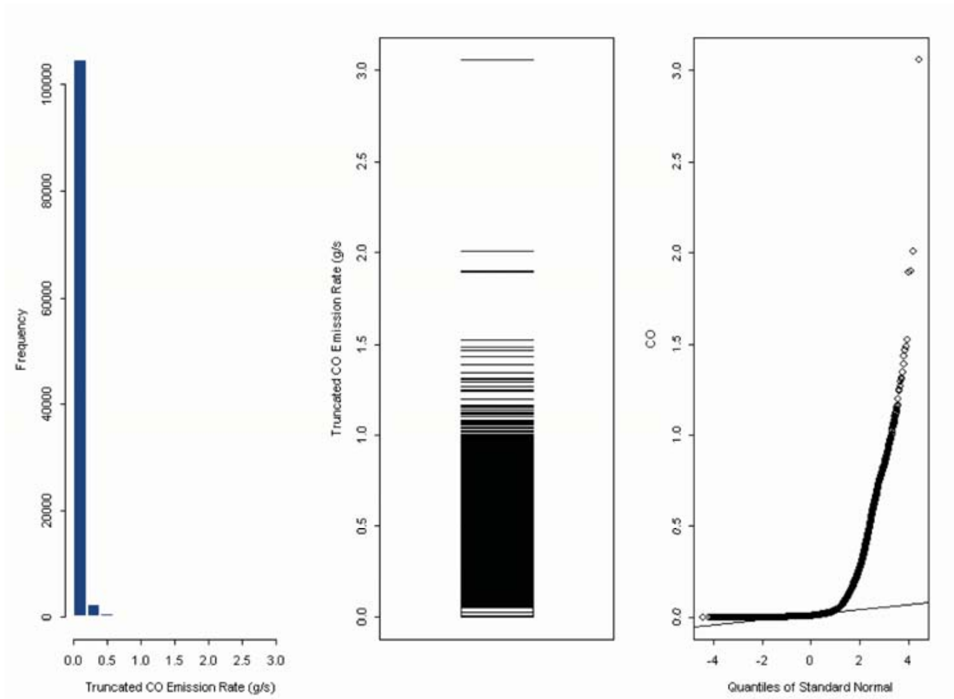


Figure 6-14 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate

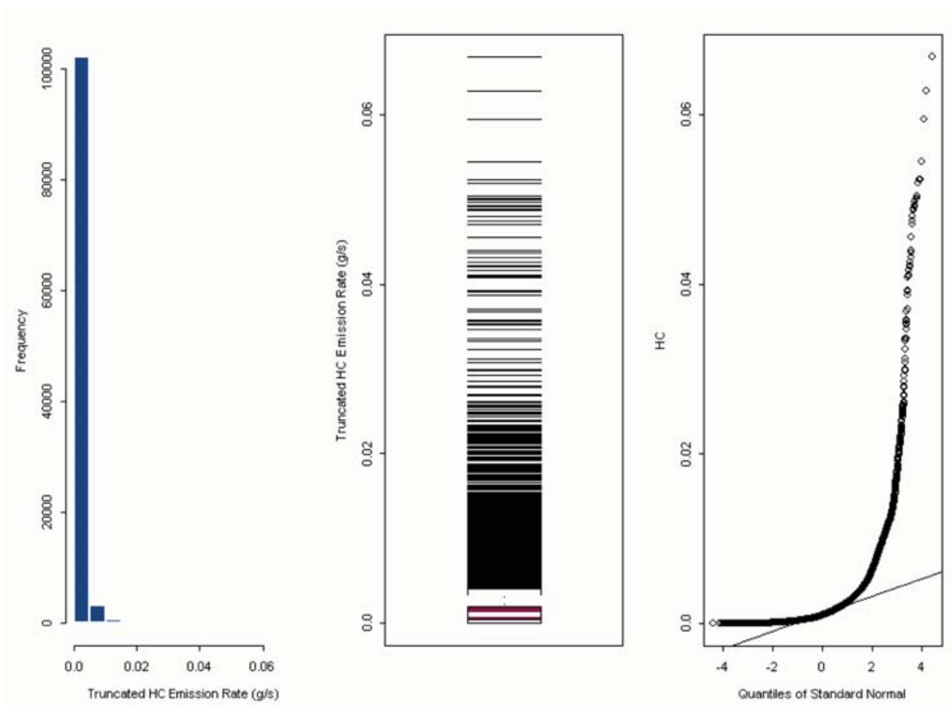


Figure 6-15 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate

These boxplots for truncated emission data show that there are some obvious outliers in the measured emissions of all three pollutants, and the histograms suggest a high degree of non-normality, also indicated in the probability plots. There is thus a need to transform the response variable to correct for this condition. Transformations are used to present data on a different scale. In modeling and statistical applications, transformations are often used to improve the compatibility of the data with assumptions underlying a modeling process, to linearize the relation between two variables whose relationship is non-linear, or to modify the range of values of a variable (Washington et al. 2003).

6.3.3 Transformation for Emissions Data

Although evidence in the literature suggests that a logarithmic transformation is most suitable for modeling motor vehicle emissions (Washington 1994; Ramamurthy et al. 1998; Fomunung 2000; Frey et al. 2002), this transformation needs to be verified through the Box-Cox procedure. The Box-Cox function in MATLAB® can automatically identify a transformation from the family of power transformations on emission data, ranging from -1.0 to 1.0. The lambdas chosen by the Box-Cox procedure are 0.22875 for truncated NO_x, -0.0648 for truncated CO, 0.14631 for truncated HC.

The Box-Cox procedure is only used to provide a guide for selecting a transformation, so overly precise results are not needed (Neter et al. 1996). It is often reasonable to use a nearby lambda value with the power transformation. The lambda values used for transformations are 1/4 for truncated NO_x, 0 for truncated CO, 0 for truncated HC. Histograms, boxplots and normal-normal plots describing transformed variable distributions are presented in Figures 6-16 to 6-18, where a great improvement is noted.

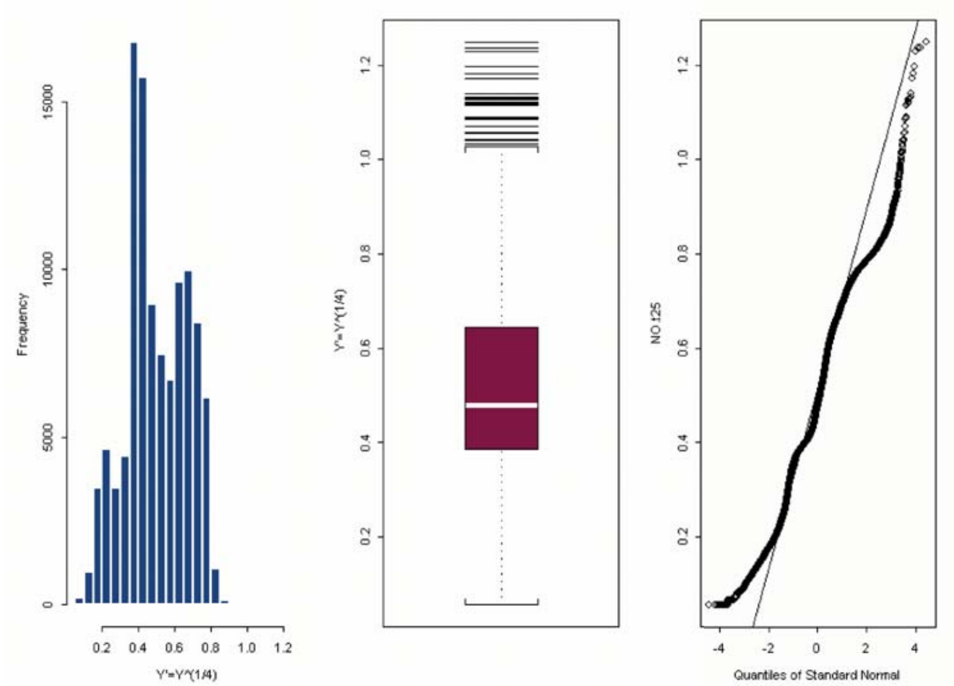


Figure 6-16 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO_x Emission Rate

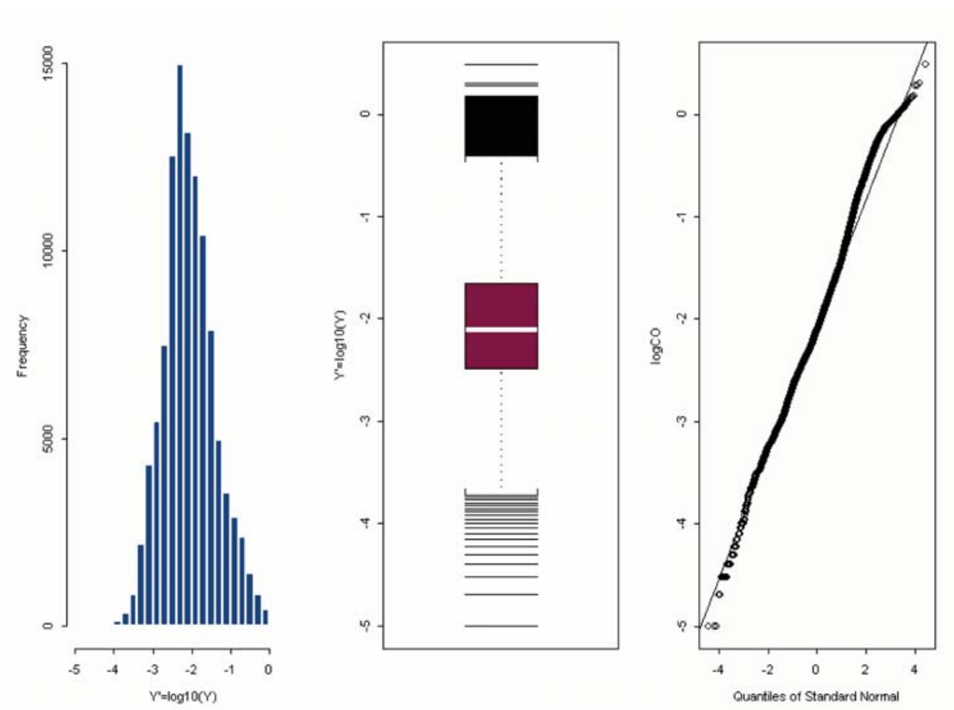


Figure 6-17 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate

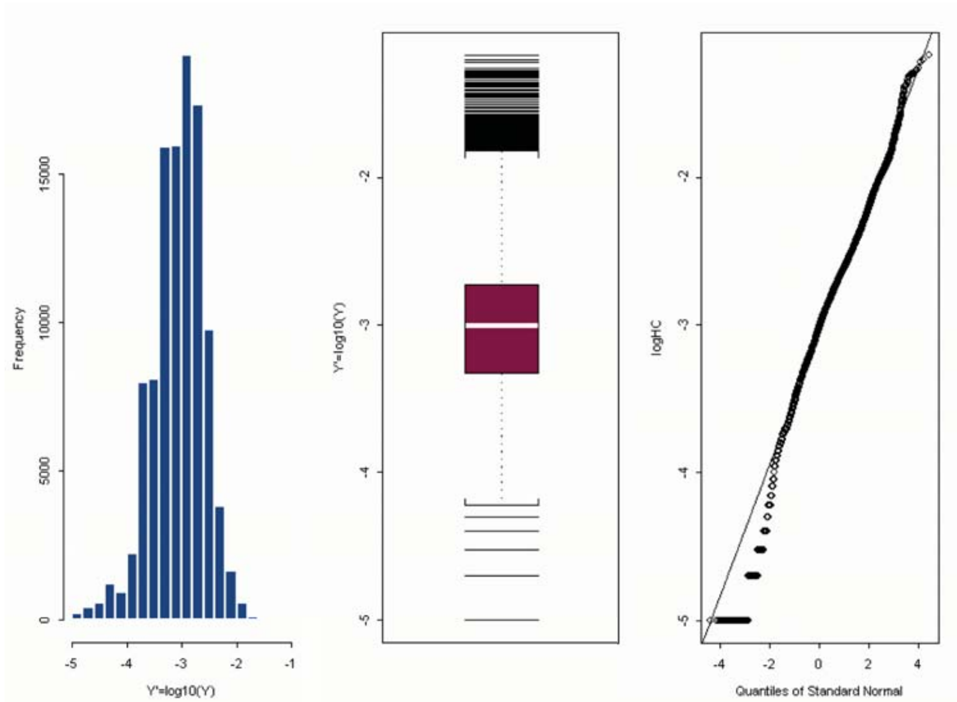


Figure 6-18 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate

Although transformations can result in improvement of a specific modeling assumption such as linearity or normality, they can often result in the violation of others. Thus, transformations must be used in an iterative fashion, with continued checking of other modeling assumptions as transformations are made. Dr. Washington suggested the comparisons should always be made on the original untransformed scale of Y when comparing statistical models and these comparisons extend to goodness of fit statistics and model validation exercises (Washington et al. 2003).

6.3.4 Identification of High Emitter

From a modeling viewpoint, it is important to accurately predict the number of 'high emitter' vehicles in the fleet (older technology, poorly maintained, or tampered vehicles that emit significantly elevated emissions relative to the fleet average under all operating conditions) and the fraction of activities that yield high emissions for normal emitting vehicles. Historic practices to identify 'high emitters' in a data set have relied on judgment to set cut points that are often indefensible from a statistical, and sometimes even practical, perspective. U.S. EPA uses five times the prevailing emission standards as the cut point across all pollutants (U.S. EPA 1993), while CARB has defined different emission regimes ranging from normal to super emitters and used different criteria for each regime (CARB 1991; Carlock 1994) (see Table 6-3) .

Table 6-3 CARB Emission Regime Definition (Carlock 1994)

Emitter Status	NO_x	CO	HC
Normal	≤ 1 standard	< 1 standard	< 1 standard
Moderate	1 to 2 standard	1 to 2 standard	1 to 2 standard
High	2 to 3 standard	2 to 6 standard	2 to 4 standard
Very High	3 to 4 standard	6 to 10 standard	5 to 9 standard
Super	> 4 standard	> 10 standard	> 9 standard

In contrast, the methodology employed in MEASURE database development at Georgia Tech is statistically based. Wolf et al. used regression tree techniques to classify vehicles into classes that behave similarly, exhibit similar technology characteristics, and exhibit similar mean emission rates under standardized testing conditions (Wolf et al. 1998). The cut points within each technology class are then defined on the basis of pre-selected percentiles of a normal distribution of the emission rates for each pollutant. The analysis by Wolf et al. specified a cut point of 97.73 percent (that is, mean + 2 standard deviations), which implies that approximately 2.27 percent of the vehicles in each technology class are high emitters.

For this research, although inter-bus variability exists in the data set, these 15 buses should be treated as one technology class because they shared the same fuel injection type, catalytic converter type, transmission type, and their model year and odometer reading were similar. Just as in Wolf's approach, the emissions value located at two standard deviations above the mean of the normalized emissions distribution is used as a cutpoint to distinguish between normal and high emission points. Theoretically, this method will consistently identify approximately 2.27 percent of the data as high emission points. That means 97.73 percent of the population should fall into the normal status. Analysis results showed that 0.33 percent of NO_x emission, 3.76 percent of CO emission, and 1.37 percent of HC emissions were identified as high emission points. After assigning those high emissions points to different buses, the distribution is shown in Table 6-4.

Table 6-4 Percent of High Emission Points by Bus

	NO_x	CO	HC
bus 360	0.02%	2.80%	5.06%
bus 361	0.32%	1.08%	0.25%
bus 363	0.06%	3.10%	0.00%
bus 364	0.04%	0.87%	7.38%
bus 372	0.00%	0.13%	1.96%
bus 375	0.69%	3.16%	0.27%
bus 377	0.00%	4.44%	0.00%
bus 379	0.67%	2.85%	1.17%
bus 380	0.52%	7.67%	0.69%
bus 381	0.10%	4.76%	0.14%
bus 382	1.14%	8.12%	0.36%
bus 383	0.88%	3.44%	1.82%
bus 384	0.50%	5.10%	1.33%
bus 385	0.55%	2.10%	0.60%
bus 386	0.20%	6.63%	0.57%
Total	0.36%	3.81%	1.38%

For each individual bus, the highest proportion is 1.14 percent for bus 382 for NO_x emissions, 8.12 percent for bus 380 for CO emissions, and 7.38 percent for bus 364 for HC emissions. No evidence from Table 6-4 suggests that there are some "high emitters" (older technology, poorly maintained, or tampered vehicles) in the data set. This conclusion makes sense since all buses were only 5 or 6 years old during the test. Another finding indicated that a small fraction of a bus's observed activity exhibited disproportionately high emissions. Activities found in the literature include hard accelerations at low speeds, moderate acceleration at high speeds, or equivalent accelerations against gravity (Fomunung 2000). Given that high emissions points make up only 0.33 percent of the data set for NO_x, 3.76 percent for CO, and 1.37 percent for HC, it is not necessary to develop two different models for normal emissions and high emissions. Based on this analysis, these 15 buses should be treated as one technology class since no high emitters were identified.

6.4 Potential Explanatory Variables

There are four main groups of parameters that affect vehicle emissions as indicated in the literature (Guensler 1993; Clark et al. 2002). These groups are: 1) vehicle characteristics, including vehicle type, make, model year, engine type, transmission type, frontal area, drag coefficient, rolling resistance, vehicle maintenance history, etc.; 2) roadway characteristics, including road grade and possibly pavement surface roughness, etc.; 3) on-road load parameters, like on-road driving trace (sec-cy-sec) or speed/acceleration profile, vehicle payload, on-road operating modes, driver behavior, etc.; and 4) environmental conditions, including humidity, ambient temperature, and ambient pressure (Feng et al. 2005; Guensler et al. 2005).

In general, emissions from HDDVs are more likely to be a function of brake-horsepower load on the engine (especially for NO_x) than emissions from light-duty gasoline vehicles, because instantaneous emissions levels of diesel engines are highly correlated with the instantaneous work output of the engine (Ramamurthy et al. 1999; Feng et al. 2005). That is, in particular, the higher the engine load, the higher emissions for NO_x. The emissions modeling framework (from which most of the items below are derived) is outlined in the Regional Applied Research Effort (RARE) report (Guensler et al. 2006). The goal of that modeling regime was to predict on-road load and then apply appropriate emission rates to the load. Most of the items outlined below are related to the amount of engine load that a vehicle will experience. Although each of the variables below is important, the values are not always available in on-road testing data (although in the future we need to make sure that these data are all collected). But, engine load in the AATA database could be used in emission rate model development for this research. Also, there are some factors, such as temperature and humidity, that may affect emission rates independent of load, or perhaps interacting with load. The model should incorporate such variables.

6.4.1 Vehicle Characteristics

Factors related to vehicle characteristics influencing heavy-duty diesel vehicle emissions which are summarized in the literature include vehicle class (i.e., weight, engine size, horsepower rating), model year, vehicle mileage, emission control system (i.e., engine exhaust aftertreatment system), transmission type, inspection and maintenance history, etc. (Guensler 1993; Clark et al. 2002).

The effect of vehicle class on emissions is significant. Five main factors that cause a vehicle to demand engine power are vehicle speed, vehicle acceleration, drive train inertial acceleration, vehicle weight, and road grade. As the required power and work performed by the vehicle increase, the amount of fuel burned to produce that power also increases, and the applicable emission rates also generally increase. Thus, emissions vary as a function of vehicle class and vehicle configuration. The higher truck classes with larger engines are heavier and, thus, typically produce more emissions. Vehicle configurations with large frontal areas and high drag coefficients will yield higher emissions when operated at higher speeds and/or accelerated at higher rates.

The concept of vehicle technology groups is to identify and track subsets of vehicles that have similar on-road load responses and similar laboratory emission rate performance. The basic premise is that vehicles in the same heavy-duty vehicle class, employing similar drive train systems, and of the same size and shape have similar load relationships. There is also an important practical consideration in establishing vehicle technology groups. Researchers need to be able to identify these vehicles in the field during traffic counting exercises.

The starting point for technology group criteria is a visual classification scheme. Yoon et al. (Yoon et al. 2004a) developed a new HDV visual classification scheme called the X-scheme based on the number of axles and gross vehicle weight ratings (GVWR) as a hybrid scheme between the FHWA truck and U.S. EPA HDV classification schemes. With field-observed HDV volumes, emissions rates estimated using the X-scheme were 34.4% and 32.5% higher for NO_x and PM, compared to using the standard U.S. EPA guidance (U.S. EPA 2004c). The X-scheme reflects vehicle composition in the field more realistically than does the standard U.S. EPA guidance (U.S. EPA 2004c), which shifted heavy-HDV volumes into light- or medium-HDV volumes 21% more frequently than the X-scheme. Figure 6-19 shows X-scheme classes and their typical figures (Yoon et al. 2004a).

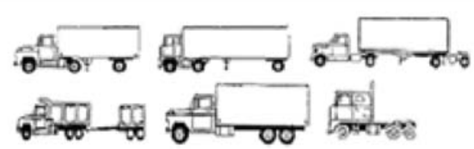
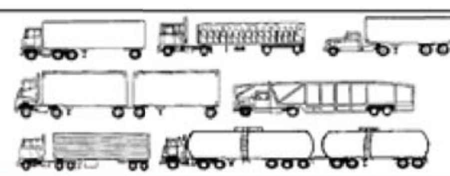
X Class	EPA Class	Typical Figures
X1	HDV2b, HDV3, HDV4, HDV5, HDV6, HDV7	
X2	HDV8a	
X3	HDV8b	

Figure 6-19 The X Classes and Typical Vehicle Configurations

Vehicle age and model year effects are accounted for because some vehicle models have much lower average emissions. Researchers from West Virginia University reported that most regulated emissions from engines produced by Detroit Diesel Corporation have declined over the years and the expected trend of decreasing emission levels with the model year of the engine is clear and consistent for PM, HC, CO and NO_x, starting with the 1990 models (Prucz et al. 2001). Information on vehicle age can be obtained from a registration database using vehicle identification numbers and truck manufacturer records. The registration database can be sorted by calendar year and show vehicles registered in the given year by model year. However, given the differences noted between field-observation fleet composition and registration data in the light-duty fleet (Granell et al. 2002), significant additional research efforts designed to model the on-road subfleet composition (classifications and model year distributions) are even more warranted for HDVs. It is also important to keep in mind that heavy-duty engines accumulate miles of travel very rapidly and that engine rebuilding is a common practice. Hence, the age of the vehicle does not necessarily equal the age of the engine. Previous field work in Atlanta indicates that on-road surveys provide better information on fleet composition (Ahanotu 1999). To refine the model, appropriate data sets that include detailed information on engine type, transmission type, etc. will be needed to appropriately subdivide the observed on-road groups and continue to develop respective emission rates. The data collection challenge in this area is daunting, but it is worthwhile to perform once to provide a library of information that can be used in a large number of modeling applications.

Vehicle weight is critical to the demand engine power that must be supplied to produce the tractive force needed to overcome inertial and drag forces and then influence vehicle emissions. NO_x emissions increase as the vehicle weight increases and this relationship does not vary much from vehicle to vehicle (Gajendran and Clark 2003). The effects of vehicle age, engine horsepower ratings, transmission type, and engine exhaust aftertreatment were also investigated in other literature (Clark et al. 2002; Feng et al. 2005).

The vast majority of heavy-duty vehicles are normal emitters, but a small percentage of vehicles are high-emitters under every operating condition, typically because they have been tampered with or they are malfunctioning (i.e., defective or mal-maintained engine sensors or actuators). As the vehicle ages, general engine wear and tear will increase emission rates moderately due to normal degradation of emission controls of properly functioning vehicles. On the other hand, as vehicles age, the probability increases that some of the vehicles will malfunction and produce significantly higher emissions (i.e., become high-emitters). Probability functions that classify vehicles within specific model years (and later, within specific statistically-derived vehicle technology groups) are currently being developed through the assessment of certification testing and various roadside emissions tests. Obtaining additional detailed sources of data for developing failure models appears to be warranted.

After engine horsepower at the output shaft has been reduced by power losses associated with fluid pressures, operation of air conditioning, and other accessory loads, there is still an additional and significant drop in available power from the engine before reaching the wheels. Power is required to overcome mechanical friction within the transmission and differential, internal working resistance in hydraulic couplings and friction of the vehicle weight on axle bearings. The combined effect of these components is parameterized as drive train efficiency. However, the more difficult and more significant component of power loss in the drive train is associated with the inertial resistance of drive train components rotational acceleration (Gillespie 1992).

A heavy-duty truck drive train is significantly more massive than its light-duty counterpart. The net effect of drive train inertial losses when operating in higher gears on the freeway may not be significant enough to be included in the model (relative to the other load-related components in the model for these heavy vehicles). However, recent studies appear to indicate very high truck emission rates (gram/second) in “creep mode” stop and start driving activities noted in ports and rail yards. Thus, high inertial loads for low gear, low speed, and acceleration operations may contribute significantly to emissions from mobile sources in freight transfer yards and therefore should not be ignored (Guensler et al. 2006).

The inertial losses are a function of a wide variety of physical drive train characteristics (transmission and differential types, component mass, etc.) and on-road operating conditions. To refine the use of inertial losses in the modal model, new drive train testing data will be designed to evaluate the inertial losses for various engine, drive shaft, differential, axle, and wheel combinations and to establish generalized drive train technology classes. Then, gear selection probability matrices for each drive train technology class and gear and final drive ratio data can be provided in lookup tables for model implementation, in place of the inertial assumptions currently employed. However, data are currently significantly lacking for development of such lookup tables.

6.4.2 Roadway Characteristics

The three basic geometric elements of a roadway are the horizontal alignment, the cross-slope or amount of super-elevation and the longitudinal profile or grade. Among them, road grade has been shown to have significant impact on engine load and vehicle emissions (Guensler 1993). Other roadway characteristics, such as lane width, are also noted to have a significant impact on the speed-acceleration profiles of heavy-duty vehicles and can therefore affect engine load (Grant et al. 1996).

6.4.3 Onroad Load Parameters

Onroad load parameters include on-road driving trace (second-by-second) or speed/acceleration profile, engine load, on-road operating modes (i.e., idling, motoring, acceleration, deceleration, and cruise), driver behavior, and so on. Vehicle speed and acceleration are integral components for the estimation of vehicle road load, and therefore engine load. Previous studies indicated that increased engine power requirements could result in the increase in NO_x emissions (Ramamurthy and Clark 1999; Feng et al. 2005). Clark et al. reported that the vehicle applications and duty cycles can have an effect on the emission produced (Clark et al. 2002). This study found that over a typical day of use for any vehicle, one that stops and then accelerates more often might produce higher distance-specific emissions, providing all else is held constant.

Passenger and freight payloads together with the vehicle tare weight contribute to the demand for power that must be supplied to produce the tractive force needed to overcome inertial and drag forces. Passenger loading functions for transit operations can be obtained through analysis of fare data or on-board passenger count programs. On the heavy-duty truck side, on-road freight weight distributions by vehicle class can be derived from roadside weigh station studies. Ahanotu conducted detailed weigh-in-motion studies in Atlanta and found that reasonable load distributions by truck class and time of day could be applied in such a modal modeling approach (Ahanotu 1999). Although additional field studies are warranted to examine the validity of the Atlanta results over time and the transferability of findings in Atlanta to other metropolitan areas (especially considering the potential variability in commodity transport, such as agricultural goods, that may occur in other areas), the modeling methodology seems appropriate.

6.4.4 Environmental Conditions

Environmental conditions under which the vehicle is operated include humidity, ambient temperature, and ambient pressure. U.S. EPA is currently conducting studies to find the effect of ambient conditions on HDDV emissions (NRC 2000). The current MOBILE6.2 model includes correction factors to account for the impact of environmental conditions on vehicle emission rates. Given the lack of compelling additional data available for analysis, it may be necessary to ignore the effects of these environmental parameters (altitude, temperature, and humidity) or simply incorporate the existing MOBILE6.2 correction factors. Preliminary analyses of the data and methods used to derive the MOBILE6.2 environmental correction factors indicate that the embedded equations in MOBIL6.2 probably need to be revisited.

6.4.5 Summary

It is impossible for modeler to include all explanatory variables identified in the literature review for model development because the explanatory variables available for model development and model validation are only a subset of potential explanatory variables identified above. Therefore, the conceptual model will only include available variables and derived variables in the data set provided.

6.5 Selection of Explanatory Variables

As mentioned earlier, available explanatory variables for transit buses are only a subset of potential explanatory variables identified. In brief, available explanatory variables can be summarized as:

- *Test information*: date, time;
- *Vehicle characteristics*: license number; model year, odometer reading, engine size, instrument configuration number;
- *Roadway characteristics*: road grade (%);
- *Onroad load parameters*: engine power (bhp), vehicle speed (mph), acceleration (mph/s);
- *Engine operating parameters*: throttle position (0 – 100%), engine oil temperature (deg F), engine oil pressure (kPa), engine warning lamp (Binary), engine coolant temperature (deg F), barometric pressure reported from ECM (kPa);
- *Environmental conditions*: ambient temperature (deg C), ambient pressure (mbar), ambient relative humidity (%), ambient absolute humidity (grains/lb air).

The most important question related to engine power is how to simulate engine power in the real world for application purposes. Georgia Institute of Technology researchers developed a transit bus engine power demand simulator (TB-EPDS), which estimates transit bus power demand for given speed, acceleration, and road grade conditions (Yoon et al. 2005a; Yoon et al. 2005b). Speed-acceleration-road grade matrices were developed from speed and location data obtained using a Georgia Tech Trip Data Collector. The researchers conclude that speed-acceleration-road grade matrices at the link level or the route level are both acceptable for regional inventory development. However, for micro-scale air quality impact analysis, link-based matrices should be employed (Yoon et al. 2005a). Although significant uncertainties still exist for inertial loss which is significant at low speeds and motoring mode with negative engine power, this research showed that using engine power as load data is possible for application purposes. Thus we concluded that engine power could be used as load data in estimated emission models.

The relationships between explanatory variables were investigated using S-Plus®. Three variables were excluded because they have only a single value for all records, and they are engine size, instrument configuration number and engine warning lamp. There are 14 explanatory variables included in correlation analysis. The correlation matrix is shown in Table 6-5.

Table 6-5 Correlation Matrix for Transit Bus Data Set

*** Correlations for data in: transitbus.data ***

	model.year	odometer	temperature	baro
model.year	1.0000000000	-0.655273106	0.047048515	0.394378106
odometer	-0.655273106	1.0000000000	0.186771499	-0.704310642
temperature	0.047048515	0.186771499	1.0000000000	-0.326938545
baro	0.394378106	-0.704310642	-0.326938545	1.0000000000
SCB.RH	0.068411842	0.343814465	0.488214011	-0.632480147
humid	0.030997734	0.39026148	0.751260451	-0.649522446
grade	-0.004241021	0.00052737	-0.005590441	0.002384338
vehicle.speed	-0.014916204	-0.062908098	-0.225478003	0.054918347
throttle.position	-0.00186824	0.009346571	-0.09113266	-0.014470281
oil.temperature	0.051759069	-0.011881827	0.042676227	-0.026744091
oil.pressure	0.050521339	-0.098442472	-0.073256993	0.034212231
coolant.temperature	0.206727241	-0.117710067	0.077114798	0.045844706
eng.bar.press	0.137781076	-0.248876183	-0.260525088	0.371021489
engine.power	-0.006066455	0.021283229	-0.059512654	-0.035718725

	SCB.RH	humid	grade	vehicle.speed
model.year	0.0684118427	0.030997734	-0.004241021	-0.014916204
odometer	0.3438144652	0.390261480	0.00052737	-0.062908098
temperature	0.4882140119	0.751260451	-0.005590441	-0.225478003
baro	-0.6324801472	-0.649522446	0.002384338	0.054918347
SCB.RH	1.0000000000	0.931879078	-0.006075112	-0.034502697
humid	0.9318790788	1.000000000	-0.006411009	-0.117870984
grade	-0.0060751123	-0.006411009	1.0000000000	0.000896568
vehicle.speed	-0.0345026977	-0.117870984	0.000896568	1.0000000000
throttle.position	0.0134235743	-0.024720165	0.020186507	0.387705398
oil.temperature	0.096018579	0.087317807	-0.007116669	0.018641433
oil.pressure	-0.0498528376	-0.077649741	0.009836954	0.567493814

coolant.temperature	0.2005559889	0.171558840	-0.014531524	0.072998199
eng.bar.press	-0.3663829274	-0.373540032	0.002132063	0.143270319
engine.power	0.0257436423	-0.003279122	0.021662091	0.303209657
acc	0.0000403711	0.003340728	0.012930076	0.000224126

	throttle.position	oil.temperature	oil.pressure
model.year	-0.001868240	0.051759069	0.050521339
odometer	0.009346571	-0.011881827	-0.098442472
temperature	-0.091132660	0.042676227	-0.073256993
baro	-0.014470281	-0.026744091	0.034212231
SCB.RH	0.013423574	0.096018570	-0.049852837
humid	-0.024720165	0.087317807	-0.077649741
grade	0.020186507	-0.007116669	0.009836954
vehicle.speed	0.387705398	0.018641433	0.567493814
throttle.position	1.000000000	0.012077329	0.681336402
oil.temperature	0.012077329	1.000000000	-0.117896787
oil.pressure	0.681336402	-0.117896787	1.000000000
coolant.temperature	0.059605193	0.335667341	-0.298083257
eng.bar.press	0.102861968	0.059886972	0.022549030
engine.power	0.959310116	0.007171781	0.656609695
acc	0.660747116	-0.004185245	0.465493435

	coolant.temperature		eng.bar.press	engine.power
model.year	0.206727200	41	0.137781076	-0.006066455
odometer	-0.117710000	67	-0.248876183	0.021283229
temperature	0.077114700	98	-0.260525088	-0.059512654
baro	0.045844700	06	0.371021489	-0.035718725
SCB.RH	0.200555900	88	-0.366382927	0.025743642
humid	0.171558800	40	-0.373540032	-0.003279122
grade	-0.014531500	24	0.002132063	0.021662091
vehicle.speed	0.072998100	99	0.143270319	0.303209657
throttle.position	0.059605100	93	0.102861968	0.959310116
oil.temperature	0.335667300	41	0.059886972	0.007171781
oil.pressure	-0.298083200	57	0.022549030	0.656609695
coolant.temperature	1.000000000	00	0.284506753	0.050584845
eng.bar.press	0.284506700	53	1.000000000	0.089702976
engine.power	0.050584800	45	0.089702976	1.000000000

All variable pairs with correlation coefficients greater than 0.5 were scrutinized and subjected to further analysis, which invariably helped in paring down the number of variables. The values in the correlation matrix show that throttle position and engine power, ambient relative humidity and ambient absolute humidity are highly correlated (higher than 0.90). Model year and odometer, odometer and barometric pressure, barometric pressure and ambient relative humidity, barometric pressure and ambient absolute humidity, ambient absolute humidity and temperature, oil pressure and throttle position, oil pressure and vehicle speed, oil pressure and engine power, throttle position and acceleration, engine power and acceleration are moderately correlated (higher than 0.50). Other pairs of variables, however, have only slight correlations.

The relationship between throttle position and engine power is shown in Figure 6-20. Since engine power is derived from percent engine load, engine torque, and engine speed, and previous studies indicated that increased engine power requirements could result in the increase in NO_x emissions (Ramamurthy and Clark 1999; Feng et al. 2005), the author retained engine power in the database.

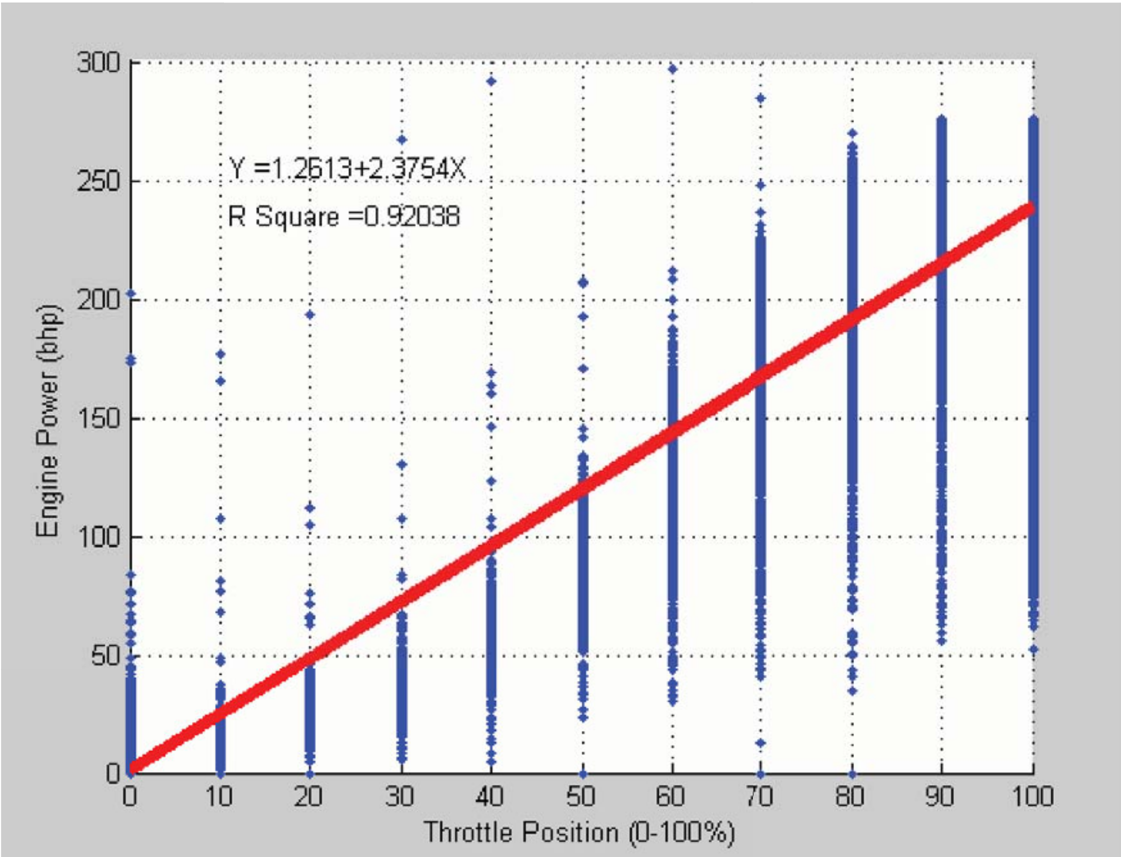


Figure 6-20 Throttle Position vs. Engine Power for Transit Bus Data Set

Ambient relative humidity and ambient absolute humidity provide the same information in two different ways, and either is enough to consider the influence of ambient humidity on emissions. The author retained ambient relative humidity in the database.

Three other findings related to the correlation matrix are:

1. All environmental characteristics, like temperature, humidity, and barometric pressure, are moderately correlated with each other (Figure 6-21), which indicates modelers should consider such relationships when developing environmental factors.

2. Engine power is correlated with not only on-road load parameters such as vehicle speed, acceleration, and road grade, but also engine operating parameters such as throttle position and engine oil pressure. Engine power in this data set is derived from measured engine speed, engine torque and percent engine load. On the other hand, engine power could be derived theoretically from vehicle speed, acceleration and road grade using an engine power demand equation. So, engine power can connect on-road modal activity with engine operating conditions at this level. This fact strengthens the importance of introducing engine power into a conceptual emissions model and to improve the ability to simulate engine power for regional inventory development.

3. Engine operating parameters, like throttle position (0 – 100%), engine oil pressure (kPa), engine oil temperature (deg F), engine coolant temperature (deg F), and barometric pressure reported from ECM (kPa), are highly or moderately related to on-road operating parameters. For example, engine power and throttle position are highly correlated, while oil pressure and vehicle speed, oil pressure and engine power, throttle position and acceleration are moderately correlated. Although engine operating parameters may have power to explain the variability of emission data, it is difficult to obtain such data in the real world for modeling purposes. These four variables are retained for further analysis of their relationships with emissions. Although these four variables will be excluded from the emission model at this time, analysis of these potential relationships may indicate a need for further research in this area.

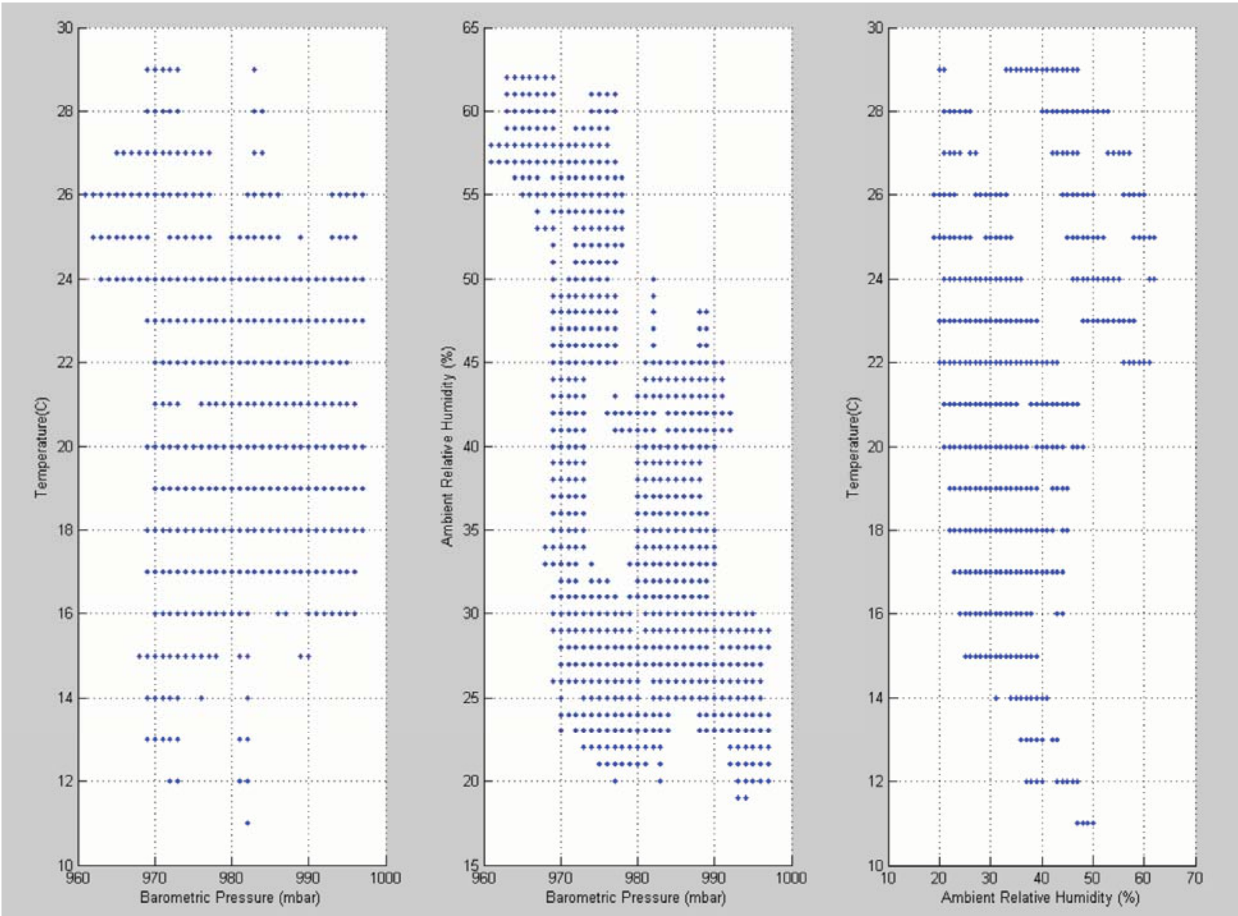


Figure 6-21 Scatter plots for environmental parameters

CHAPTER 7

7. MODAL ACTIVITY DEFINITIONS DEVELOPMENT

7.1 Overview of Current Modal Activity Definitions

Current research suggests that vehicle emission rates are highly correlated with modal vehicle activity. Modal activity is a vehicle activity characterized by cruise, idle, acceleration or deceleration operation. Consequently, a modal approach to transportation-related air quality modeling is becoming widely accepted as more accurate in making realistic estimates of mobile source contribution to local and regional air quality. Research at Georgia Tech has clearly identified that modal operation is a better indicator of emission rates than average speed (Bachman 1998). The analysis of emissions with respect to driving modes, also referred to as modal emissions, has been done in several recent researches (Barth et al. 1996; Bachman 1998; Fomunung et al. 1999; Frey et al. 2002; Nam 2003; Barth et al. 2004). These studies indicated that driving modes might have the ability to explain a significant portion of variability of emission data. Usually, driving can be divided into four modes: acceleration, deceleration, cruise, and idle. But driving mode definitions in literature were somewhat arbitrary. To define the driving modes or choose more reasonable definitions for the proposed modal emissions model, current driving mode definitions used in different modal emission models need to be investigated first.

MEASURE's Definitions

Researchers at Georgia Tech developed the MEASURE model in 1998 (Guensler et al. 1998). This model was developed from more than 13,000 laboratory tests conducted by the EPA and CARB using standardized test cycle conditions and alternative cycles (Bachman 1998). Modal activities variables were introduced into the MEASURE model as follows: *acceleration* (mph/sec), *deceleration* (mph/sec), *cruise* (mph) and percent in idle time. In addition, two surrogate variables were also developed, *inertial power surrogate* (IPS) (mph^2/s), which was defined as acceleration times velocity and *drag power surrogate* (DPS) (mph^3/s), which was defined as acceleration times velocity squared. Within each mode, several 'cut points', or threshold values, were specified and used to create several categories. In total, six threshold values were defined for acceleration, three for deceleration, five for cruise modes, seven for IPS, and seven for DPS. Modal activity surrogate variables were added as percent of cycle time spend in specified operating conditions (Fomunung et al. 1999).

NCSU's Definitions

Dr. Frey at NCSU defined four modes of operation (idle, acceleration, deceleration, and cruise), for U.S. EPA's MOVES' model in 2001 (Frey and Zheng 2001; Frey et al. 2002). The following description is directly cited from his report (Frey et al. 2002).

> Idle is defined as based upon zero speed and zero acceleration. The acceleration mode includes several considerations. First, the vehicle must be moving and increasing in speed. Therefore, speed must be greater than zero and the acceleration must be greater than zero. However, vehicle speed can vary slightly during events that would typically be judged as cruising. Therefore, in most instances, the acceleration mode is based upon a minimum acceleration of 2 mph/sec. However, in some cases, a vehicle may accelerate slowly. Therefore, if the vehicle has had a sustained acceleration rate averaging at least 1 mph/sec for at least three seconds or more, that is also considered acceleration. Deceleration is defined in a similar manner as acceleration, except that the criteria for deceleration are based upon negative acceleration rates. All other events not classified as idle, acceleration, or deceleration, are classified as cruising. Thus, cruising is approximately steady speed driving but some drifting of speed is allowed.

Physical Emission Rate Estimator's (PERE's) Definitions

Dr. Nam developed his definitions when he introduced his Physical Emission Rate Estimator (PERE) model in 2003 (Nam 2003). Idle is defined as speed less than 2 mph. Acceleration mode is based on acceleration rate greater than 1 mph/sec. However, deceleration is based on deceleration rate less than -0.2 mph/sec. Other events are classified as cruise mode and the acceleration range is between -0.2 mph/sec and 1 mph/sec. Nam also mentioned in his report that the definition of cruise (based only on acceleration) will change depending on the speed in future studies.

Summary

Current driving mode definitions related to modal emission models are all significantly different from each other. NCSU used one absolute critical value, 2 mph/sec, for acceleration and deceleration mode. However, PERE chose two different critical values, 1 mph/sec and -0.2 mph/sec, for acceleration and deceleration mode individually. The critical values, 2 mph/sec, 1 mph/sec, or 0.2 mph/sec, were chosen somewhat arbitrarily. MEASURE used several threshold values to add modal activity surrogate variables. Table 7-1 summarizes these modal activity definitions.

Table 7-1 Comparison of Modal Activity Definition

	MEASURE	NCSU	PERE
Idle	Speed=0, Acc=0	Speed=0, Acc=0	Speed<2
Acceleration	Acc>6,Acc>5,Acc>4, Acc>3,Acc>2,Acc>1	Acc>2 or Acc>1 for three seconds	Acc>1
Deceleration	Acc<-3,Acc<-2, Acc<-1	Acc<-2 or Acc<-1 for three seconds	Acc<-0.2
Cruise	Speed>70, Speed>60, Speed>50, Speed>40, Speed>30	Other events	-0.2<Acc<1

Note: Unit for speed is mph, unit for acceleration is mph/sec.

7.2 Proposed Modal Activity Definitions and Validation

Although the current mode definitions could all explain some variability in different emission data sets (Barth et al. 1996; Bachman 1998; Fomunung et al. 1999; Frey et al. 2002; Nam 2003; Barth et al. 2004), they differ significantly from each other. Determining whether to accept current definitions or develop new definitions is therefore a challenge.

MEASURE's definitions were developed based on cycle test data and modal activity surrogate variables were added as percent of cycle time spent in specified operating conditions. Obviously, this definition is not suitable for second-by-second data. PERE's definition could not assign all data into appropriate modes. Idle mode was defined as zero speed and zero acceleration in NCSU's definitions. Although idle mode is defined theoretically as zero speed and zero acceleration, idle mode could not be defined in this manner without considering unavoidable measurement error and measurement noise. Based on this analysis, it seems more reasonable to develop new definitions for this proposed modal emission model, where such definitions can be derived through empirical analysis of the data. In fact, the definition of modal activity is dependant on the available speed/acceleration data and data quality. For example, a lack of zero speed records does not mean that there is no idle activity in the data set.

The initial proposed modal activity definitions were defined as follows:

- Idle is defined as based on speeds less than 2.5 mph and absolute acceleration less than 0.5 mph/sec.

- Acceleration mode is based upon a minimum acceleration of 0.5 mph/sec.

- Deceleration is defined in a manner similar to acceleration, except that the criteria for deceleration are based upon negative acceleration rates.

- All other events not classified as idle, acceleration, or deceleration, are classified as cruise.

At the same time, several different critical values were chosen to examine the reasonableness of the proposed criteria. Four different mode definitions using different critical values are shown in Table 7-2.

Table 7-2 Four Different Mode Definitions and Modal Variables

	Idle	Acceleration	Deceleration	Cruise
Definition 1	Speed $\leq$ 2.5 & abs(acc) $\leq$ 0.5	Acc > 0.5	Acc < -0.5	Other
Definition 2	Speed $\leq$ 2.5 & abs(acc) $\leq$ 1	Acc > 1	Acc < -1	Other
Definition 3	Speed $\leq$ 2.5 & abs(acc) $\leq$ 1.5	Acc > 1.5	Acc < -1.5	Other
Definition 4	Speed $\leq$ 2.5 & abs(acc) $\leq$ 2	Acc > 2	Acc < -2	Other

Note: Unit for speed is mph, unit for acceleration is mph/sec.

A program was written in MATLAB™ to determine the driving mode for second-by-second data and estimate the average value of emissions for each of the driving modes. At the same time, average modal emission rates were estimated for each mode based on different modal activity definitions in Table 7-2. Figures 7-1 to 7-3 present a comparison of average modal emission rates for different pollutants (NO_x, CO, and HC).

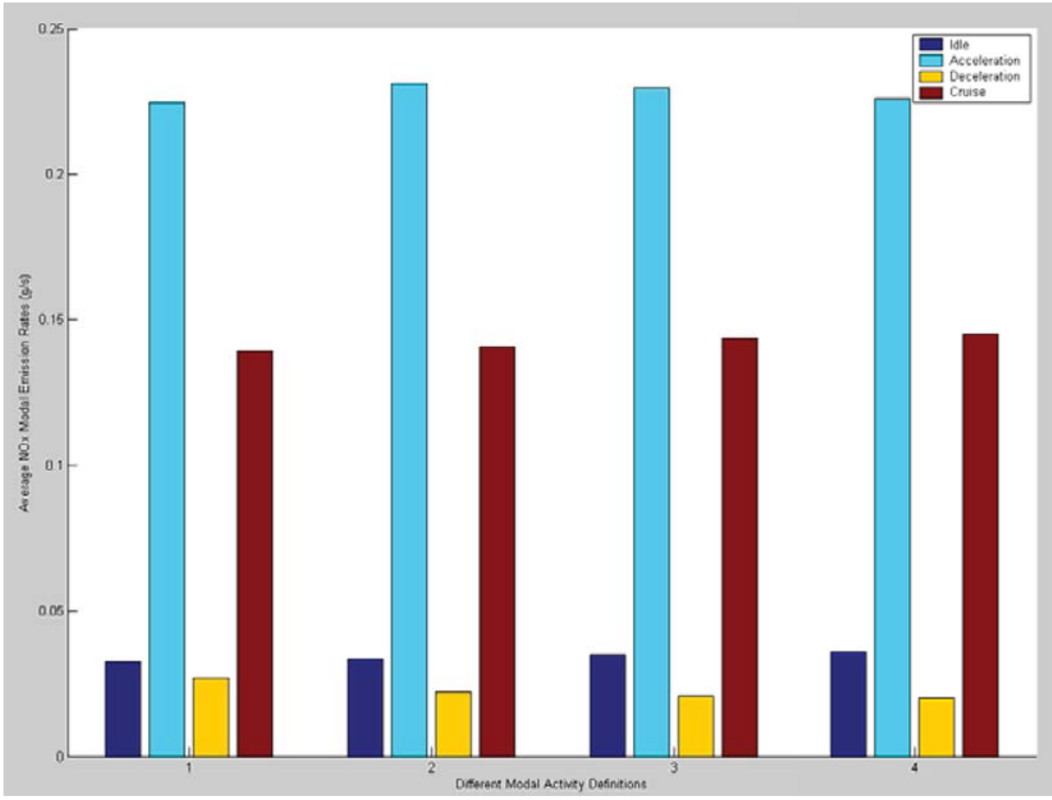


Figure 7-1 Average NO_x Modal Emission Rates for Different Activity Definitions

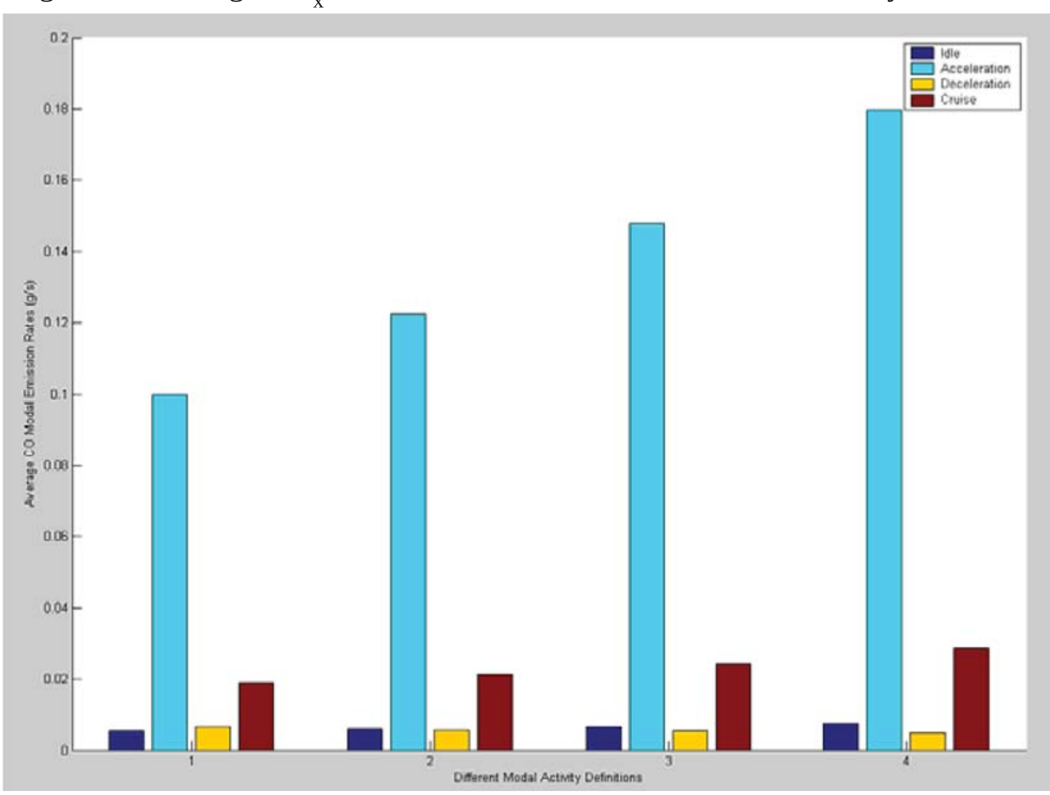


Figure 7-2 Average CO Modal Emission Rates for Different Activity Definitions

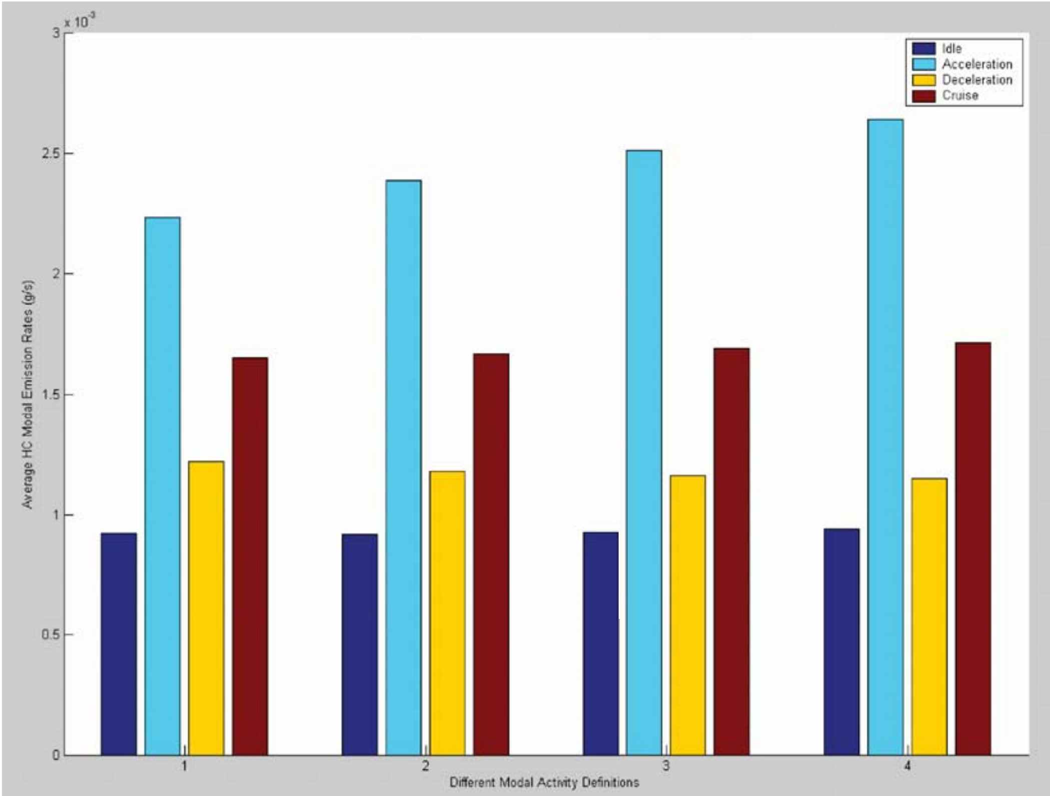


Figure 7-3 Average HC Modal Emission Rates for Different Activity Definitions

These four different modal activity definitions show a kind of consistent pattern. The average emissions during the acceleration mode are significantly higher than any other driving mode for all of the pollutants. The average emission rate during deceleration mode is the lowest of the four modes for NO_x and CO emissions while the average emission rate during idle mode is the lowest of the four modes for HC emissions. The average cruising emission rate is typically higher than the average idling and decelerating emission rate, except for CO emission in definitions 3 and 4.

To assess whether the average modal emission rates are statistically significantly different from each other, two-sample tests were estimated for each pair. Lilliefors tests for goodness of fit to a normal distribution were first used for each mode based on different modal activity definitions. The results show that all of them reject the null hypothesis of normal distribution at 5% level. A Kolmogorov-Simirnov two-sample test was chosen to take place of the *t*-test because the assumption of normal distribution was questionable. The Kolmogorov-Smirnov two-sample test is a test of the null hypothesis that two independent samples have been drawn from the same population (or from populations with the same distribution). The test uses the maximal difference between cumulative frequency distributions of two samples as the test statistic. Results of the Kolmogorov-Smirnov two-sample tests are presented in Table 7-3 in terms of p-values where

“Acc” represents acceleration mode while “Dec” represents deceleration mode. The cases where the p-value is less than 0.05 indicate that the distributions are different at the 5% level. All p-values for 72 possible pairwise comparisons are lower than 0.05, indicating that the distributions for these pairs are statistically different from each other.

Table 7-3 Results for Pairwise Comparison for Modal Average Estimates In Terms of P-value

		Idle-Acc	Idle-Dec	Idle-Cruise	Acc-Dec	Acc-Cruise	Dec-Cruise
Definiton1	NO_x	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0
Definiton2	NO_x	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0
Definiton3	NO_x	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0
Definiton4	NO_x	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0

The modal emission analysis results suggest that all four mode definitions proposed in Table 7-2 appear reasonable. These modal definitions allow some explanation of differences in emissions based upon driving mode, as revealed by the fact that the modal emission distributions differ from each other. A further step is taken here to see which mode definition would be identified as the most appropriate definition by utilizing HTBR technique. For each definition, three dummy variables are added to represent idle, acceleration, and deceleration mode. The regression trees are developed between emission data and these three dummy variables for each definition are shown in Figures 7-4 to 7-6. The sensitivity test results based on these regression trees for NO_x, CO, and HC are summarized in Table 7-4.

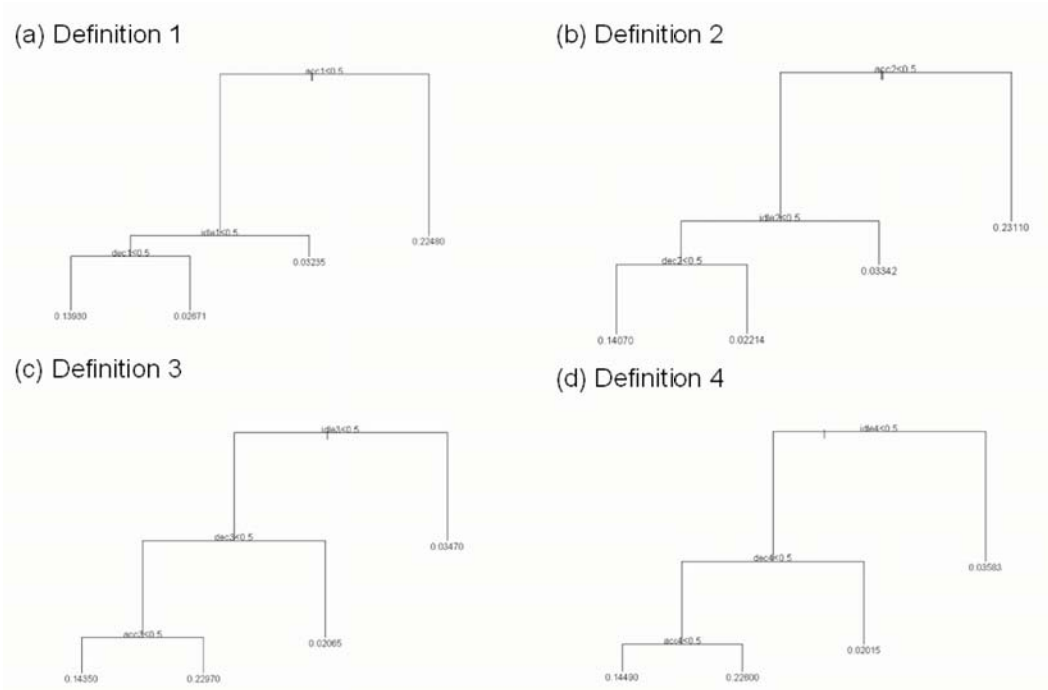


Figure 7-4 HTBR Regression Tree Result for NO_x Emission Rate

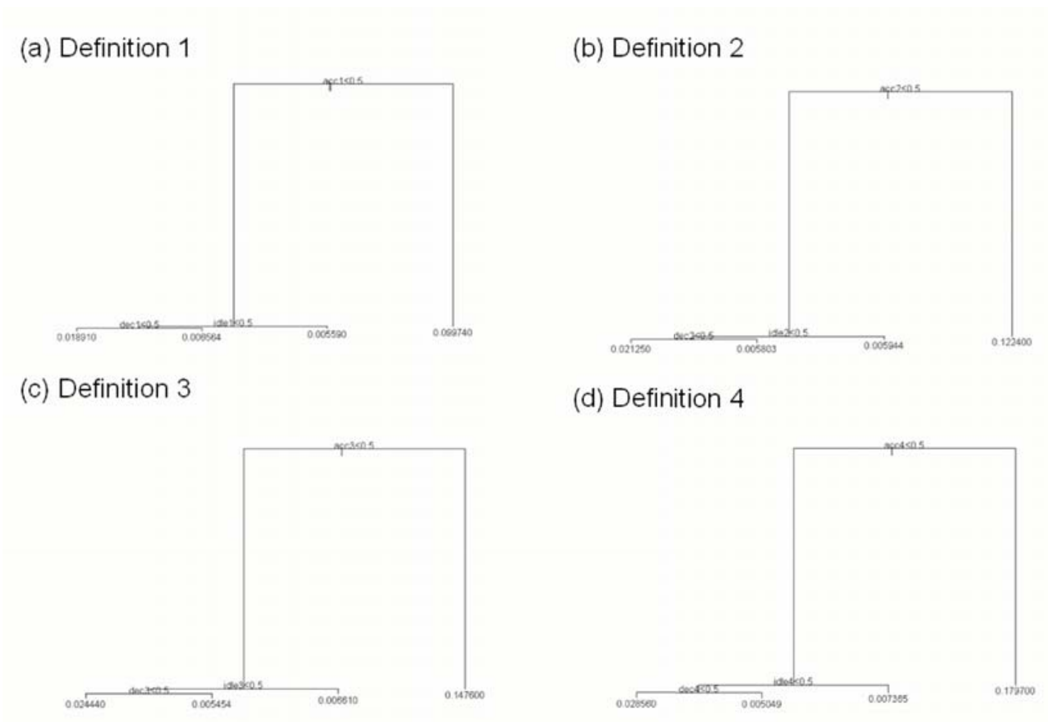


Figure 7-5 HTBR Regression Tree Result for CO Emission Rate

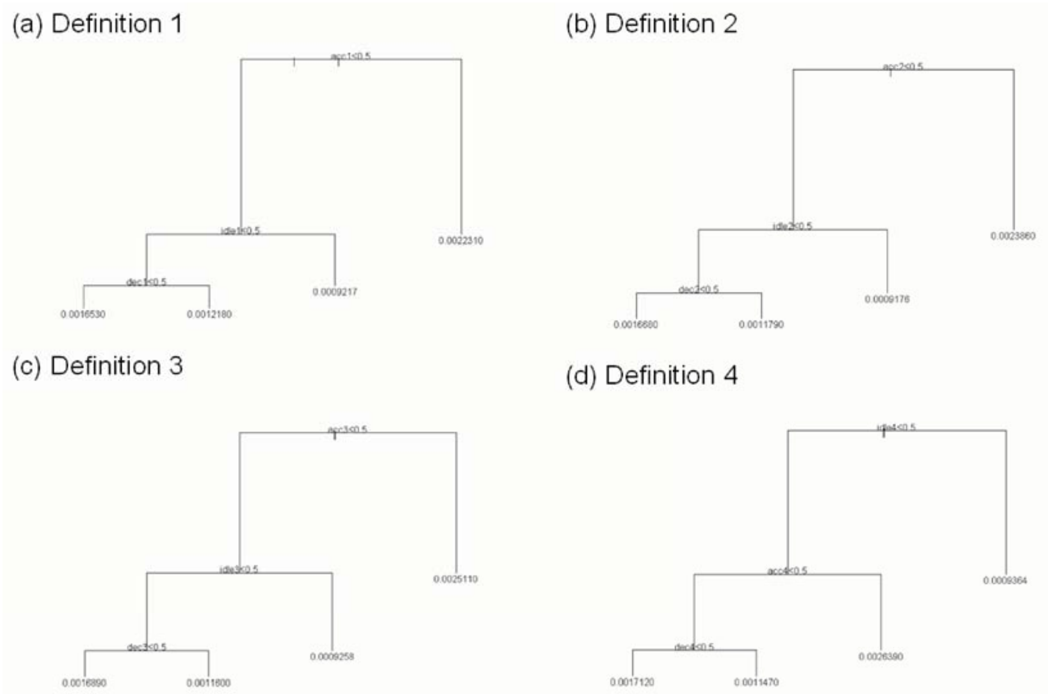


Figure 7-6 HTBR Regression Tree Result for HC Emission Rate

Table 7-4 Sensitivity Test Results for Four Mode Definition

NO_x	Mode	Number	Deviance	Mean ER	Residual Mean Deviance
		105976	1435.00	0.10680	
Definition 1					0.006967 = 738.3 / 106000
	Idle	29541	11.04	0.03235	
	Acceleration	25931	320.90	0.22480	
	Deceleration	22242	41.32	0.02671	
	Cruise	28262	365.10	0.13930	
Definition 2					0.007658 = 811.5/106000
	Idle	31064	16.05	0.03342	
	Acceleration	18894	206.50	0.23110	
	Deceleration	16644	21.14	0.02214	
	Cruise	39374	567.80	0.14070	
Definition 3					0.00856 = 907.1 / 106000
	Idle	32010	23.07	0.03470	
	Acceleration	13417	130.50	0.2297	
	Deceleration	12768	14.27	0.02065	
	Cruise	47781	739.30	0.14350	

NO_x	Mode	Number	Deviance	Mean ER	Residual Mean Deviance
Definition 4					0.009397 = 995.8 / 106000
	Idle	32717	30.240	0.03583	
	Acceleration	8719	77.150	0.22600	
	Deceleration	9452	9.191	0.02015	
	Cruise	55088	879.200	0.14490	
CO					
		105765	771.300	0.032370	
Definition 1					0.005795 = 612.9 / 105800
	Idle	29287	2.166	0.005590	
	Acceleration	25866	559.400	0.099740	
	Deceleration	22456	3.903	0.006564	
	Cruise	28156	47.380	0.018910	
Definition 2					0.005486283 = 580.2 / 105800
	Idle	30764	4.185	0.005944	
	Acceleration	18864	484.900	0.122400	
	Deceleration	16919	2.410	0.005803	
	Cruise	39218	88.710	0.021250	
Definition 3					0.005293 = 559.8 / 105800
	Idle	31691	9.131	0.006610	
	Acceleration	13402	410.100	0.147600	
	Deceleration	13035	1.861	0.005454	
	Cruise	47637	138.700	0.024440	
Definition 4					0.005239 = 554 / 105800
	Idle	32375	15.5200	0.007365	
	Acceleration	8712	339.1000	0.179700	
	Deceleration	9681	0.7047	0.005049	
	Cruise	54997	198.7000	0.028560	
HC					
		103405	0.40270	0.0014960	
Definition 1					3.648e-006 = 0.3772 / 103400
	Idle	28780	0.09337	0.0009217	
	Acceleration	25122	0.09143	0.0022310	
	Deceleration	22287	0.07644	0.0012180	
	Cruise	27216	0.11600	0.0016530	
Definition 2					3.629e-006 = 0.3752 / 103400
	Idle	30250	0.09492	0.0009176	
	Acceleration	18330	0.06668	0.0023860	
	Deceleration	16805	0.05355	0.0011790	
	Cruise	38020	0.16010	0.0016680	

NO_x	Mode	Number	Deviance	Mean ER	Residual Mean Deviance
Definition 3					3.636e-006 = 0.376 / 103400
	Idle	31157	0.09651	0.0009258	
	Acceleration	12999	0.04355	0.0025110	
	Deceleration	12970	0.04256	0.0011600	
	Cruise	46279	0.19330	0.0016890	
Definition 4					3.656e-006 = 0.378 / 103400
	Idle	31849	0.09835	0.0009364	
	Acceleration	8443	0.02944	0.0026390	
	Deceleration	9613	0.03257	0.0011470	
	Cruise	53500	0.21760	0.0017120	

7.3 Conclusions

Comparison of modal average estimates shows that the average modal emission rates are statistically different from each other for three different pollutants. HTBR regression tree results demonstrate that all four definitions can work well to divide the database. Comparisons of residual mean deviance indicate that definition 1 has the smallest residual mean deviance for NO_x (definition 4 for CO and definition 2 for HC). However, differences were small. At this time, it is difficult to choose one definition for three pollutants based just on sensitivity analysis results in this chapter. The analysis results in this section indicate that driving mode definition could not be transferred directly from one research study to another research study. A better approach would be to test several different critical values and obtain the most suitable definition instead of testing only one definition developed from other research. For this research, more analysis will be performed in the chapters that follow to develop the most suitable driving mode definitions.

CHAPTER 8

8. IDLE MODE DEVELOPMENT

In Chapter 7, the concept of driving modes was introduced and several sensitivity tests (comparison of modal average estimates, comparison of HTBR regression tree results, and comparison of residual mean deviance) were performed for four different mode definitions. Based on sensitivity analysis results, it is difficult to choose one definition for three pollutants at this moment. More analysis will be performed next to develop the most suitable driving mode definition. This chapter will focus on developing the suitable definition for idle mode.

Theoretically, idle mode is usually defined as zero speed and zero acceleration. In real world data collection efforts, this definition must be refined due to the presence of speed measurement error. In this research, idle mode will be defined by speed and acceleration. The critical value could not be deduced directly from previous research. It is better to test several critical values statistically and identify the most suitable idle definition.

8.1 Critical Value for Speed in Idle Mode

Three critical values were tested to get the appropriate critical value for speed in defining idle activity. Figures 8-1 to 8-3 illustrate engine power vs. emission rates for three pollutants for three critical speed values: 1 mph, 2.5 mph, and 5 mph. Figure 8-4 compares engine power distributions for these three critical values. Because engine power distributions for three pollutants exhibit similar patterns, only NO_x emissions are shown in Figure 8-4. Tables 8-1 and 8-2 provide the engine power distribution for these three critical values in two ways: by number and percentage.

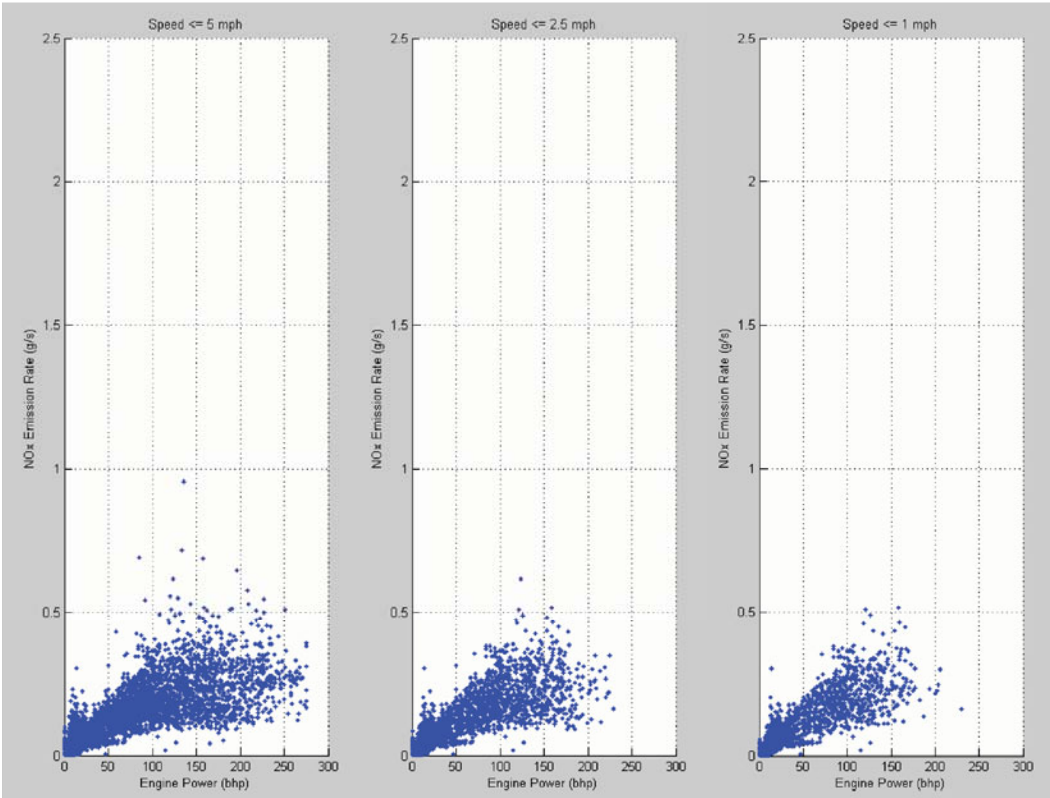


Figure 8-1 Engine Power vs. NO_x Emission Rate for Three Critical Values

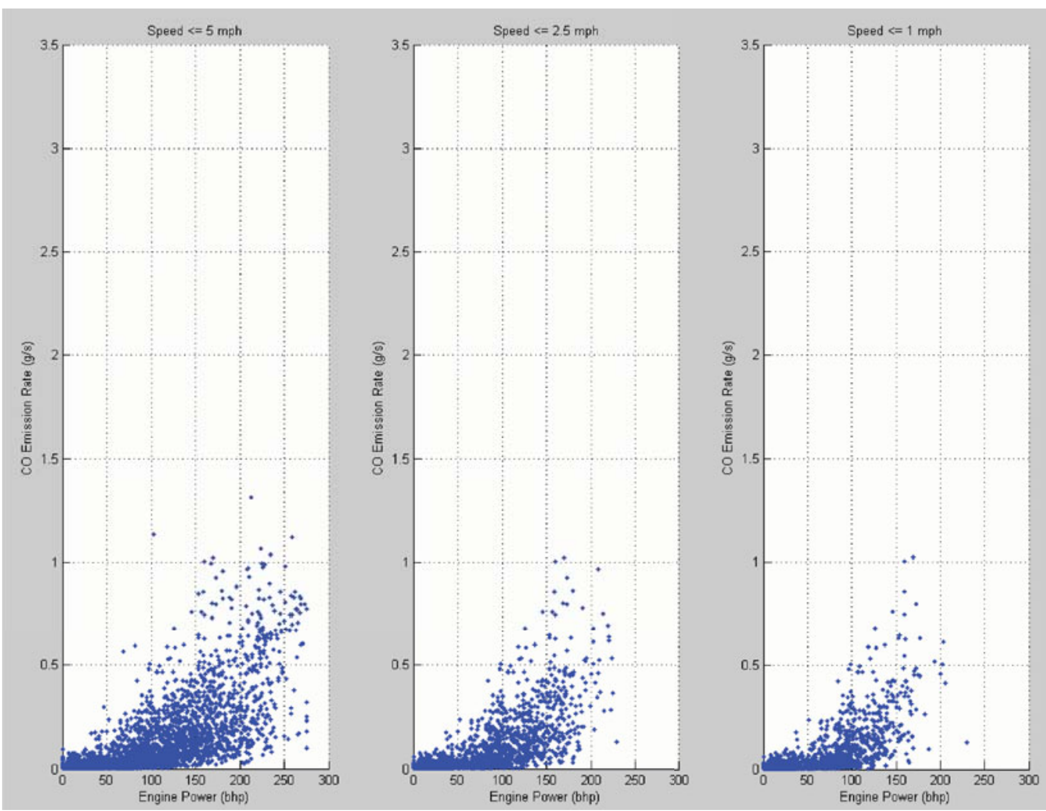


Figure 8-2 Engine Power vs. CO Emission Rate for Three Critical Values

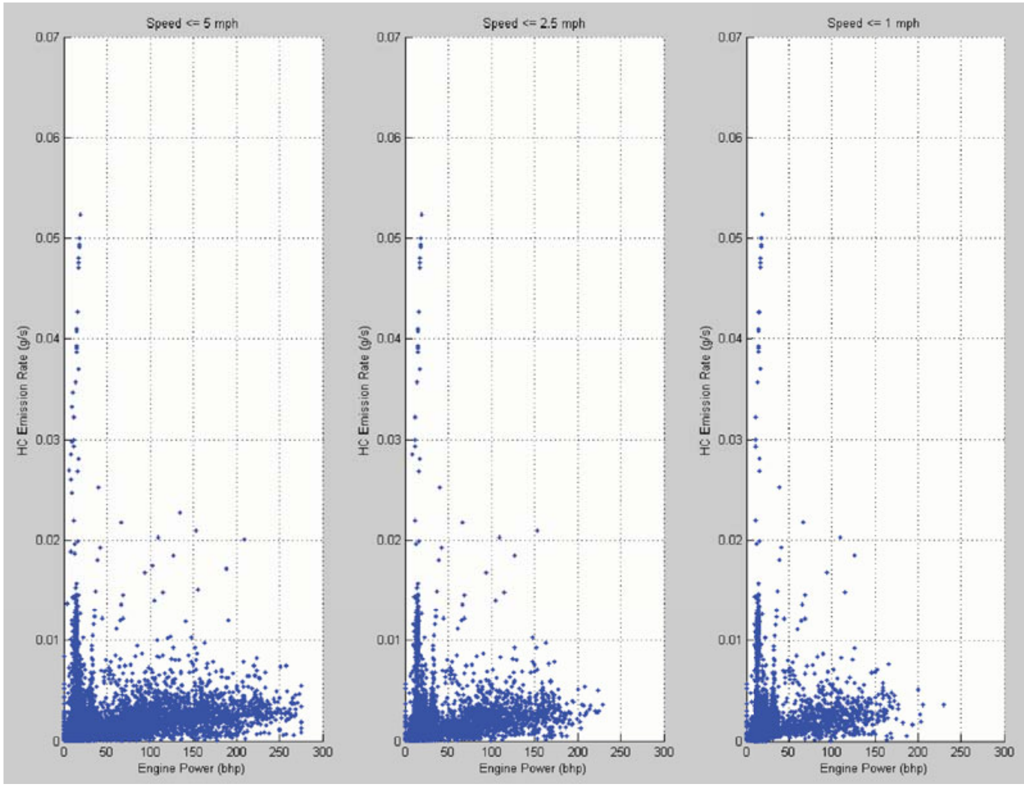


Figure 8-3 Engine Power vs. HC Emission Rate for Three Critical Values

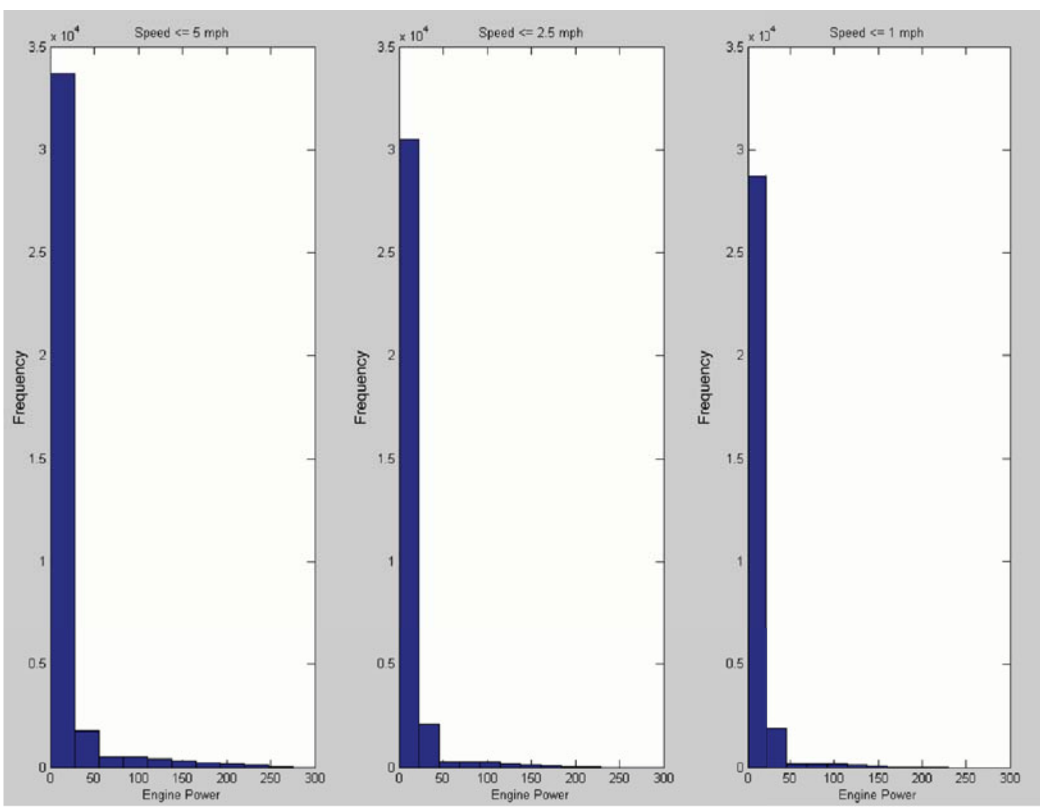


Figure 8-4 Engine Power Distribution for Three Critical Values based on NO_x Emissions

Table 8-1 Engine Power Distribution for Three Critical Values for Three Pollutants

Speed	Pollutant	Engine Power (brake horsepower (bhp)					
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	Total
≤ 5 mph	NO_x	31631	2272	1323	152	2348	37726
	CO	31258	2269	1316	149	2342	37334
	HC	30737	2264	1321	147	2284	36753
≤ 2.5mph	NO_x	29222	2098	1196	83	1143	33742
	CO	28880	2096	1189	81	1139	33385
	HC	28373	2093	1194	80	1106	32846
≤ 1 mph	NO_x	27516	2011	1100	51	700	31378
	CO	27217	2010	1093	51	699	31070
	HC	26713	2007	1099	48	680	30547

Table 8-2 Percentage of Engine Power Distribution for Three Critical Values for Three Pollutants

Speed	Pollutant	Engine Power (brake horsepower (bhp)					
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	Total
≤ 5 mph	NO_x	83.84%	6.02%	3.51%	0.40%	6.22%	100%
	CO	83.73%	6.08%	3.52%	0.40%	6.27%	100%
	HC	83.63%	6.16%	3.59%	0.40%	6.21%	100%
≤ 2.5mph	NO_x	86.60%	6.22%	3.54%	0.25%	3.39%	100%
	CO	86.51%	6.28%	3.56%	0.24%	3.41%	100%
	HC	86.38%	6.37%	3.64%	0.24%	3.37%	100%
≤ 1 mph	NO_x	87.69%	6.41%	3.51%	0.16%	2.23%	100%
	CO	87.60%	6.47%	3.52%	0.16%	2.25%	100%
	HC	87.45%	6.57%	3.60%	0.16%	2.23%	100%

Based on the analysis above, a critical value of 5 mph includes more data points with higher engine power (>50 bhp) than 2.5 mph and 1 mph. However, there is no large difference for engine power distributions between 2.5 mph and 1 mph. These two critical values for speed will be tested further with different acceleration values in the next section. The results will be used to make a final decision with regards to deceleration mode.

8.2 Critical Value for Acceleration in Idle Mode

After setting the critical value for speed, the next step is to determine a critical value for acceleration. In total, four options were tested.

- Option 1: speed ≤ 2.5 mph and absolute acceleration ≤ 2 mph/s

- Option 2: speed ≤ 2.5 mph and absolute acceleration ≤ 1 mph/s
- Option 3: speed ≤ 1 mph and absolute acceleration ≤ 2 mph/s
- Option 4: speed ≤ 1 mph and absolute acceleration ≤ 1 mph/s

Using the same method as outlined in the previous section, Figures 8-5 to 8-7 illustrate engine power vs. emission rates for three pollutants for four options above. Figure 8-8 compares engine power distribution for data falling into these four options. Because engine power distributions for three pollutants exhibit a similar pattern, only NO_x emissions are shown in Figure 8-8. Tables 8-3 and 8-4 provide the engine power distribution for four options in two ways: by number and percentage.

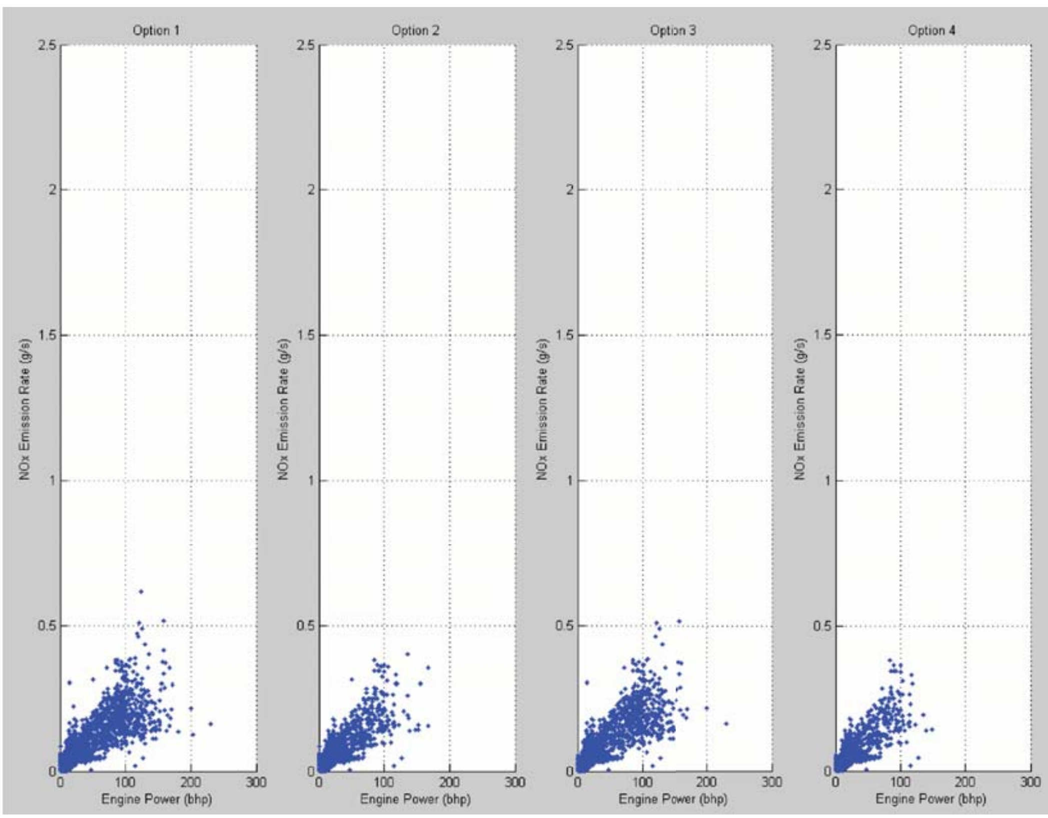


Figure 8-5 Engine Power vs. NO_x Emission Rate for Four Options

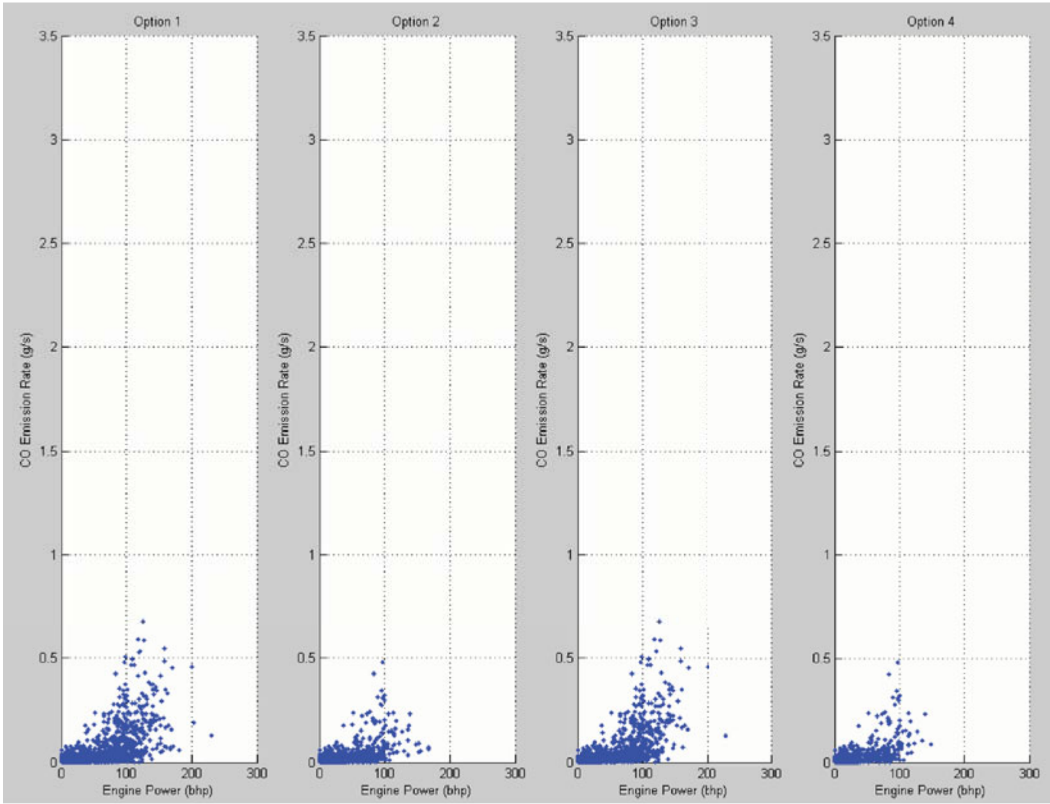


Figure 8-6 Engine Power vs. CO Emission Rate for Four Options

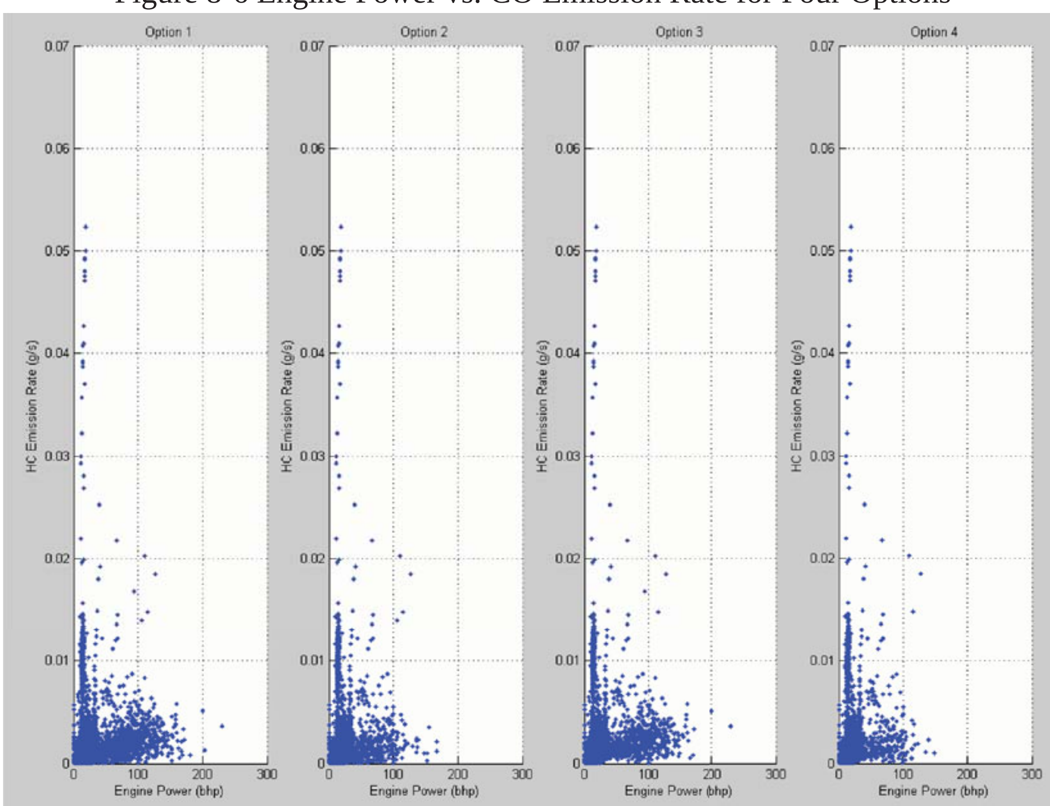


Figure 8-7 Engine Power vs. HC Emission Rate for Four Options

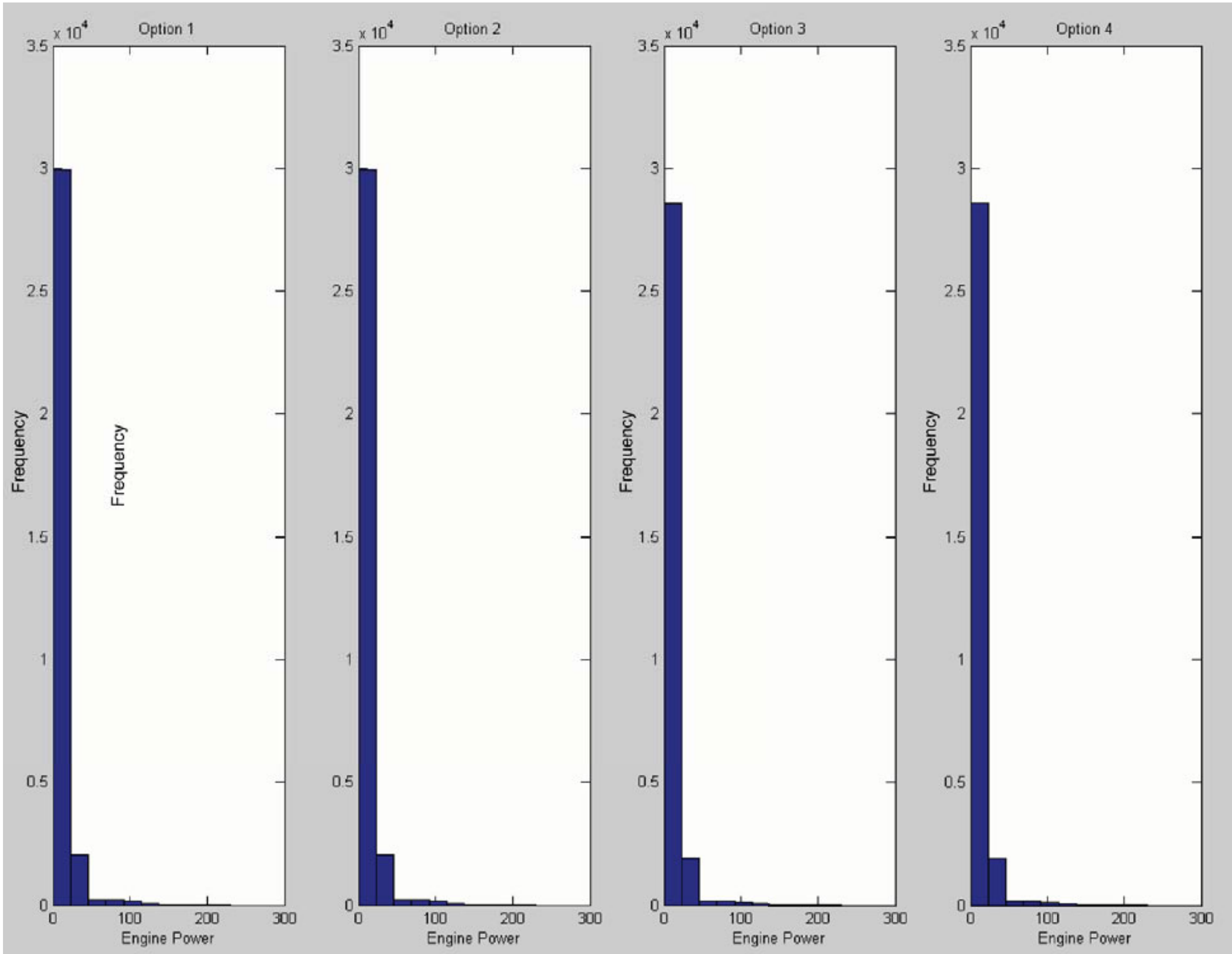


Figure 8-8 Engine Power Distribution for Four Options based on NO_x Emission Rates

Table 8-3 Engine Power Distribution for Four Options for Three Pollutants

	Pollutants	Engine Power (brake horsepower (bhp)) [0 20)	[20 30)	[30 40)	[40 50)	≥ 50	Total
Option 1	NO_x	28694	2075	1177	78	693	32717
	CO	28366	2073	1170	76	690	32375
	HC	27855	2070	1175	75	674	31849
Option 2	NO_x	27571	2030	1120	53	290	31064
	CO	27284	2028	1114	51	287	30764
	HC	26771	2026	1119	51	283	30250
Option 3	NO_x	27367	1999	1091	50	527	31034
	CO	27071	1998	1084	50	526	30729
	HC	26569	1995	1090	47	512	30213
Option 4	NO_x	26719	1969	1057	34	205	29984
	CO	26446	1968	1051	34	204	29703
	HC	25944	1966	1056	32	198	29196

Table 8-4 Percentage of Engine Power Distribution for Three Critical Values for Three Pollutants

	Pollutants	Engine Power (brake horsepower (bhp)					
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	Total
Option 1	NO_x	87.70%	6.34%	3.60%	0.24%	2.12%	100.00%
	CO	87.62%	6.40%	3.61%	0.23%	2.13%	100.00%
	HC	87.46%	6.50%	3.69%	0.24%	2.12%	100.00%
Option 2	NO_x	88.76%	6.53%	3.61%	0.17%	0.93%	100.00%
	CO	88.69%	6.59%	3.62%	0.17%	0.93%	100.00%
	HC	88.50%	6.70%	3.70%	0.17%	0.94%	100.00%
Option 3	NO_x	88.18%	6.44%	3.52%	0.16%	1.70%	100.00%
	CO	88.10%	6.50%	3.53%	0.16%	1.71%	100.00%
	HC	87.94%	6.60%	3.61%	0.16%	1.69%	100.00%
Option 4	NO_x	89.11%	6.57%	3.53%	0.11%	0.68%	100.00%
	CO	89.03%	6.63%	3.54%	0.11%	0.69%	100.00%
	HC	88.86%	6.73%	3.62%	0.11%	0.68%	100.00%

Based on the above analysis, data falling into option 2 and option 4 contain fewer data points with higher engine power (>50 bhp) than data falling into option 1 and option 3. But a large difference is not observerd in the engine power distribution for data falling into option 2 and option 4. Based upon these results, the idle mode is defined as speed ≤ 2.5 mph and absolute acceleration ≤ 1 mph/s.

8.3 Emission Rate Distribution by Bus in Idle Mode

After defining “speed ≤ 2.5 mph and absolute acceleration ≤ 1 mph/s” as idle mode, emission rate histograms for each of the three pollutants for idle operations are presented in Figure 8-9. Figure 8-9 shows significant skewness for all three pollutants for idle mode. Inter-bus response variability for idle mode operations is illustrated in Figures 8-10 to 8-12 using median and mean of NOx, CO, and HC emission rates. Table 8-5 presents the same information in tabular form. The difference between median and mean is also an indicator of skewness.

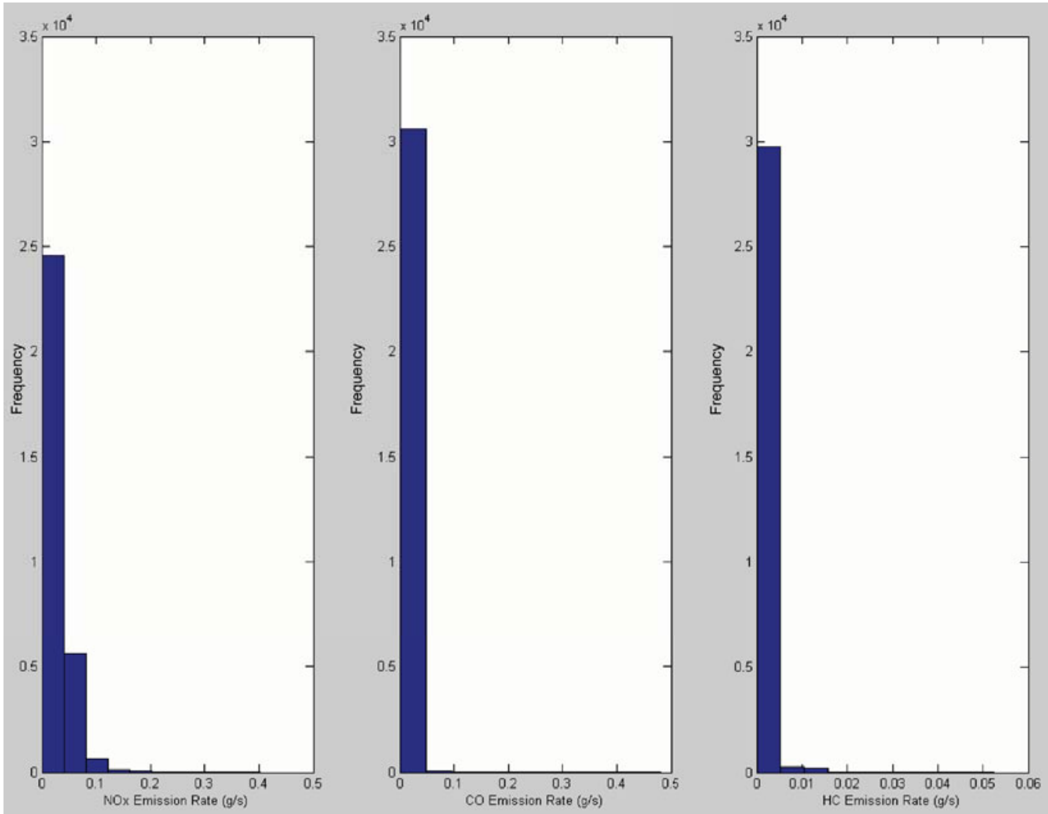


Figure 8-9 Histograms of Three Pollutants for Idle Mode

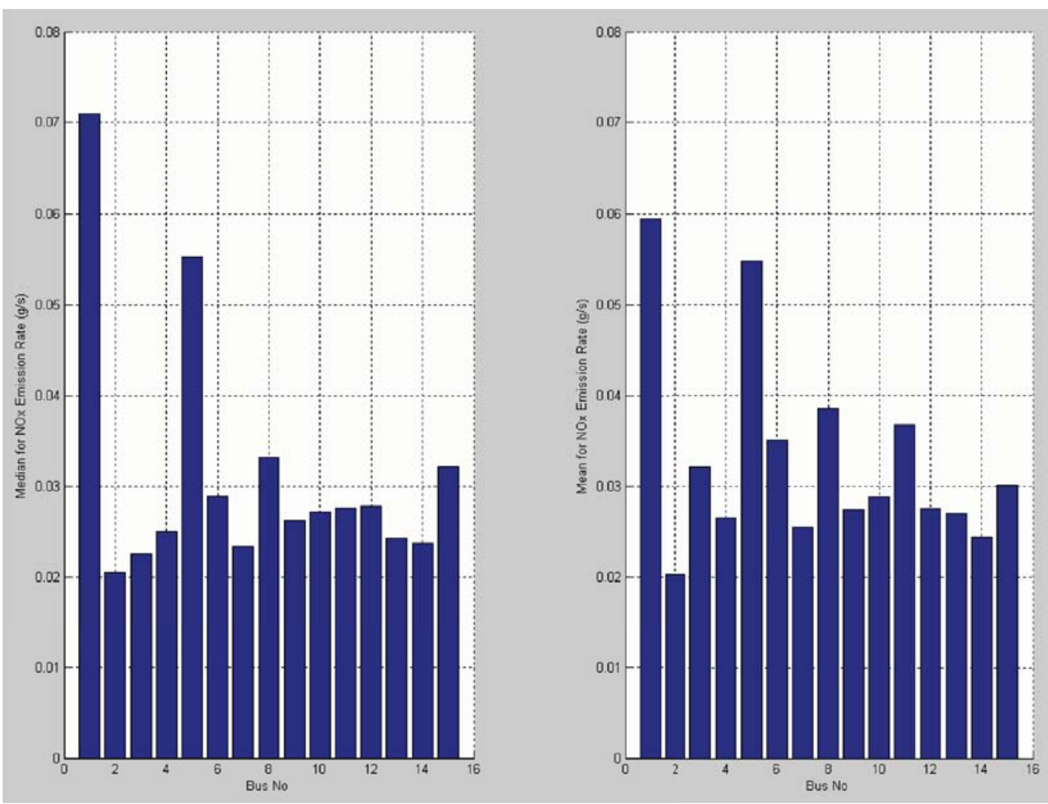


Figure 8-10 Median and Mean of NO_x Emission Rates in Idle Mode by Bus

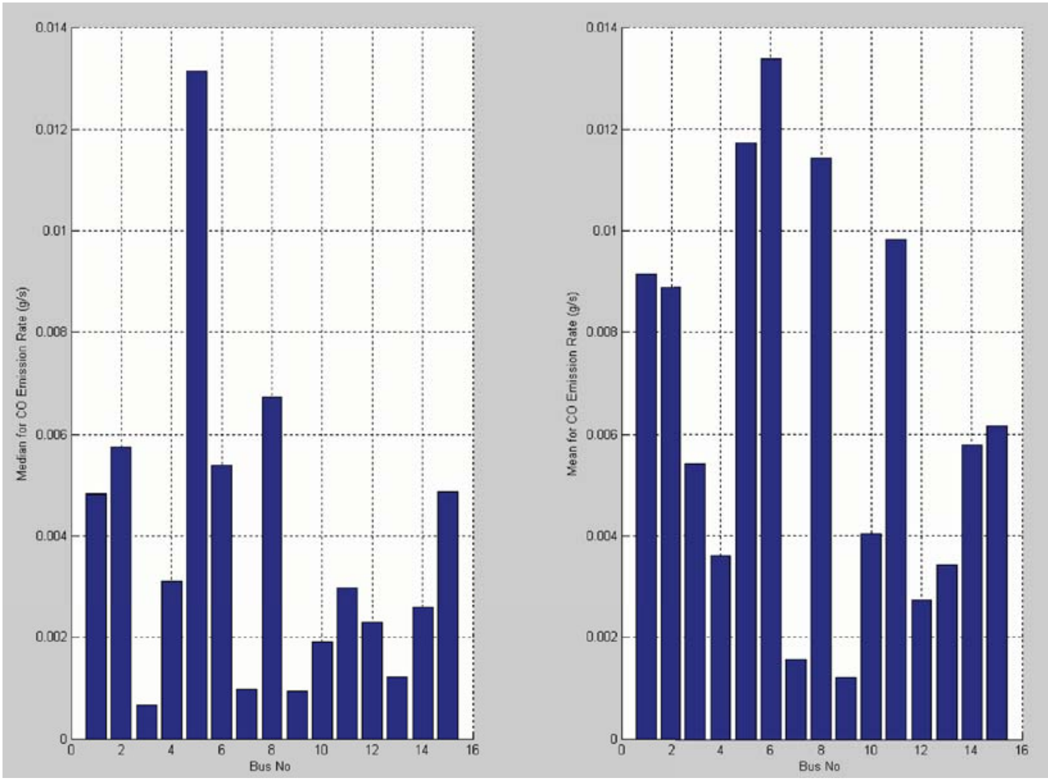


Figure 8-11 Median and Mean of CO Emission Rates in Idle Mode by Bus

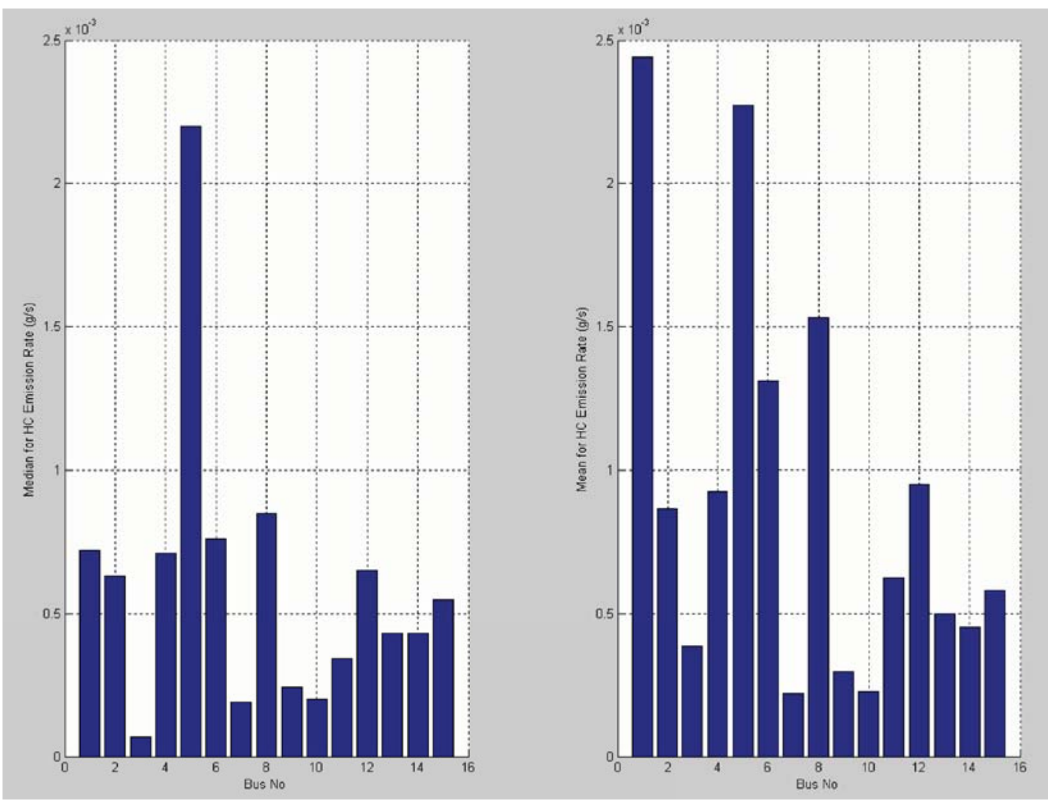


Figure 8-12 Median and Mean of HC Emission Rates in Idle Mode by Bus

Table 8-5 Median, and Mean of Three Pollutants in Idle Mode by Bus

	NO_x		CO		HC	
Bus ID	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**
Bus 360	0.071020	0.059444	0.004830	0.009145	0.00072	0.002441
Bus 361	0.020455	0.020216	0.005740	0.008895	0.00063	0.000865
Bus 363	0.022555	0.032140	0.000670	0.005408	0.00007	0.000385
Bus 364	0.025050	0.026480	0.003110	0.003601	0.00071	0.000927
Bus 372	0.055210	0.054766	0.013150	0.011739	0.00220	0.002272
Bus 375	0.028880	0.035050	0.005390	0.013385	0.00076	0.001311
Bus 377	0.023370	0.025393	0.000960	0.001572	0.00019	0.000219
Bus 379	0.033210	0.038500	0.006730	0.011425	0.00085	0.001531
Bus 380	0.026200	0.027371	0.000930	0.001218	0.00024	0.000298
Bus 381	0.027115	0.028768	0.001915	0.004044	0.00020	0.000228
Bus 382	0.027605	0.036734	0.002980	0.009836	0.00034	0.000624
Bus 383	0.027790	0.027520	0.002290	0.002736	0.00065	0.000950
Bus 384	0.024210	0.026982	0.001205	0.003428	0.00043	0.000498
Bus 385	0.023750	0.024339	0.002590	0.005782	0.00043	0.000453
Bus 386	0.032140	0.030031	0.004860	0.006155	0.00055	0.000579

Figures 8-10 to 8-12 and Table 8-5 illustrate that bus 372 has the largest median and the second largest mean for CO and HC emissions, and the second largest median and the second largest mean for NO_x emissions. The activity of bus 372 in terms of distribution of engine power by bus was compared to that of other buses in an effort to identify why the emission rates were significantly higher than for other buses. Table 8-6 and Figure 8-13 show that bus 372 has higher min (2^{nd}), 1^{st} quartile (2^{nd}), median (1^{st}), and 3^{rd} quartile (2^{nd}) engine power compared to the other 14 buses. Engine power in idle mode may include cooling fan, air compressor, air conditioner, and alternator loads (Clark et al. 2005). Considering test buses and engines are similar in many ways, this difference might be caused by variability across the engines, or may be associated with unrecorded air conditioner use. In analyzing the database, the modeler could not identify a contribution of air conditioner to engine power in idle mode. So, model development will include these data but readers should be cautioned that the noted variability is an indication that significant numbers of vehicles may need to be tested in the future if such inter-engine differences are significant in the fleet. In addition, the role of air conditioning usage on engine load in transit buses warrants additional future research.

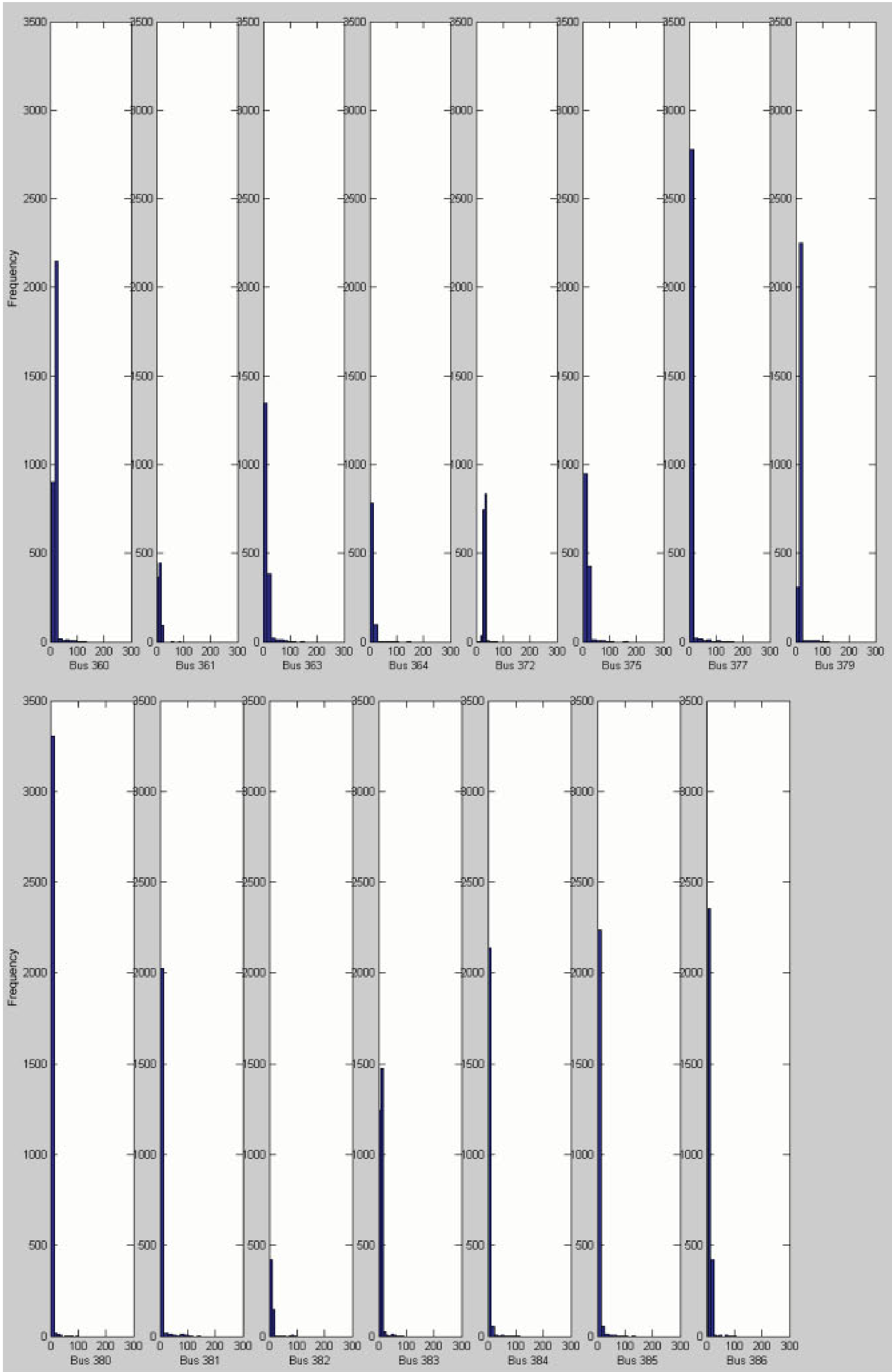


Figure 8-13 Histograms of Engine Power in Idle Mode by Bus

Table 8-6 Engine Power Distribution in Idle Mode by Bus

Bus ID	Min	1st Quartile	Median	3rd Quartile	Max
Bus 360	3.92	15.36	18.7	19.83	135.43
Bus 361	0	5.35	12.52	13.83	89.47
Bus 363	0	13.1	13.34	15.16	152.94
Bus 364	0	13.18	13.85	14.99	154.51
Bus 372	0	26.44	31.84	33.10	79.08
Bus 375	0	12.52	13.81	18.08	167.72
Bus 377	0	8.5	9.17	9.85	166.86
Bus 379	0	15.86	17.15	19.42	126.64
Bus 380	2.67	7.85	8.49	9.17	100.99
Bus 381	0	8.7	10.49	11.17	148.28
Bus 382	0	7.35	8.52	13.89	99.04
Bus 383	0	7.16	10.03	12.5	91.86
Bus 384	0	6.01	7.34	8.51	117.39
Bus 385	0	4.53	7.19	8.51	139.05
Bus 386	4.68	9.18	13.33	14.46	105.44

8.4 Discussions

8.4.1 High HC Emissions

Figure 8-7 shows that there are some high HC emissions in idle mode. Based on definitions of “speed ≤ 2.5 mph and absolute acceleration ≤ 1 mph/s”, 388/30250=1.28% of data points in idle mode for HC are high emissions. These high emissions were noted in the HC emissions data, not in NO_x and CO. All high HC emissions have been coded as high-idle to determine if they are related to any other parameters. Tree analysis could be used for this screening analysis. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM pressure, and other parameters, no specific operating parameters related to these high-idle emissions were identified.

On the other hand, regression tree analysis results by bus and trip are presented in Figure 8-14. The left figure shows that these high HC emissions occurred in bus 360 and 372 while the right figure shows that these high HC emissions happened in bus 360 trip 4 and bus 372 trip 1. Even for HC emissions, Figure 8-14 shows that these high emissions are not a common situation in idle mode. There are 1529 idle segments in total for 15 buses, but most of these high HC emissions came just from three idle segments. These three idle segments are: bus 360 trip 4 idle segment 1 (130 seconds), bus 360 trip 4 idle segment 38 (516 seconds) and bus 372 trip 1 idle segment 1 (500 seconds). More specifically, bus 360 trip 4 idle segment 1 contains 102 high HC emissions, bus 360 trip 4 idle segment 38 contains 264 high HC emissions, while bus 372 trip 1 idle slots contain 13 high HC emissions. Figures 8-15 to 8-17 illustrate time series plots for HC for these three idle segments while vehicle speed, engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature and ECM pressure are presented, too. These figures do not include NO_x and CO because NO_x and CO do not show such patterns as these three idle segments for HC. These three idle segments contain 379 high HC emissions in total. Thus about 98% of high emissions came from three idle segments only. Exclusion of these three idle segments based on all current information is difficult. The modeler prefers to keep these data since these outliers might reflect variability in the real world. However, future data collection efforts should seek to identify the causes of such events.

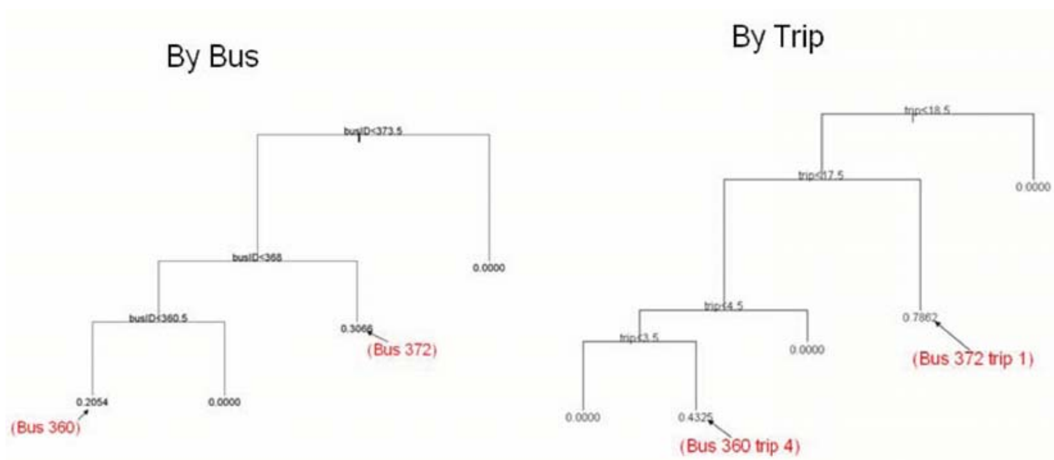


Figure 8-14 Tree Analysis Results for High HC Emission Rates by Bus and Trip

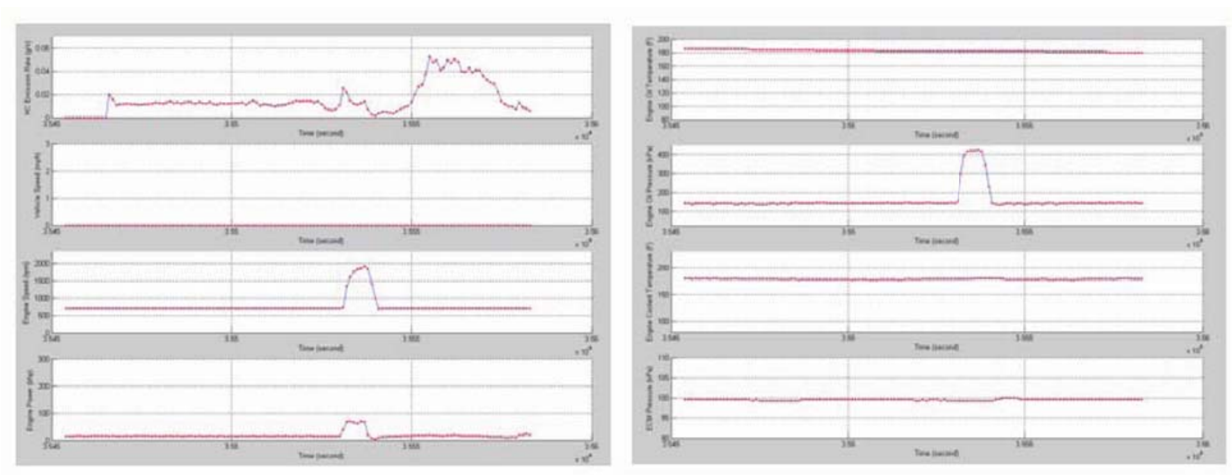


Figure 8-15 Time Series Plot for Bus 360 Trip 4 Idle Segment 1 (130 Seconds)

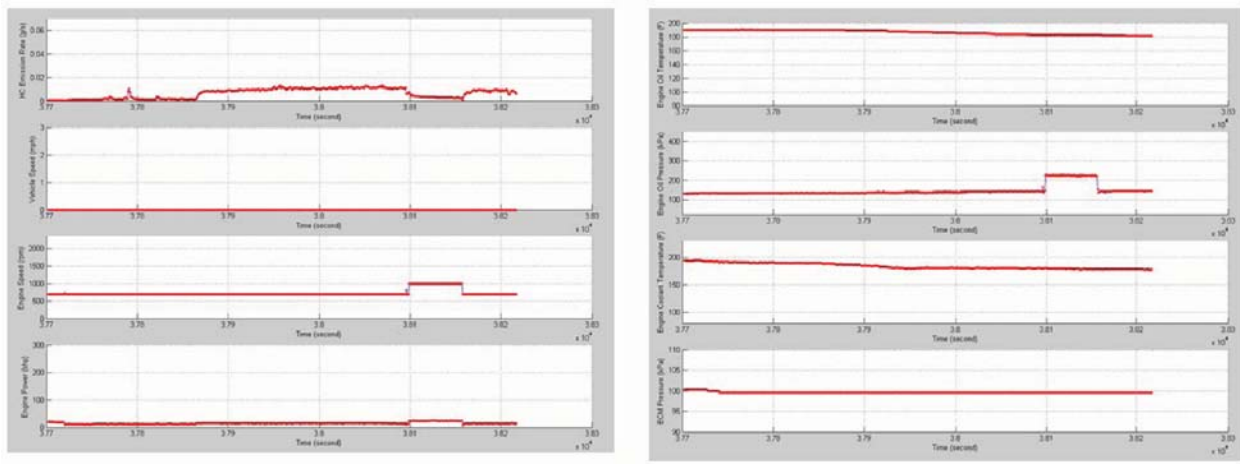


Figure 8-16 Time Series Plot for Bus 360 Trip 4 Idle Segment 38 (516 Seconds)

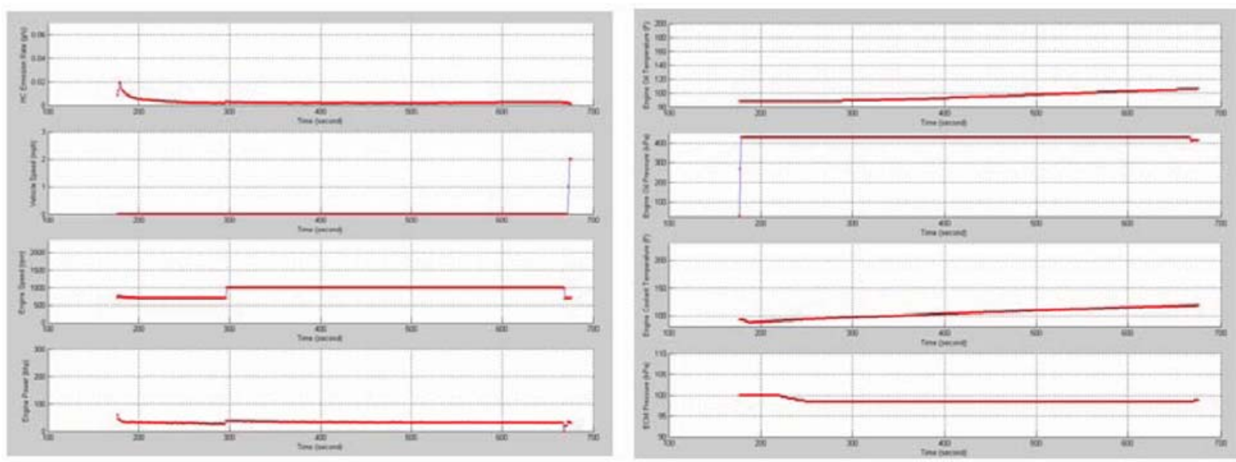


Figure 8-17 Time Series Plot for Bus 372 Trip 1 Idle Segment 1 (500 Seconds)

8.4.2 High Engine Operating Parameters

Figure 8-15 shows that engine speed once jumped to about 2000 rpm during bus 360 trip 4 idle segment 1, while corresponding engine power and engine oil pressure jumped, too. This jump lasted only 9 seconds. There are several reasons which might be responsible for this jump. Possibly bus 360 moved slowly from one location to another location while the GPS failed to detect the movement. Other explanations might be that the engine experienced a computer or sensor problem. This kind of jump, higher engine speeds (about 2000 rpm) accompanied by higher engine power and engine oil pressure in idle mode, did occur in the real world. The jump shown in Figure 8-16 was not such an occurrence since engine speed was only about 1000 rpm during that jump. After screening the whole dataset, another example of a jump is shown in Figure 8-18. The jump in bus 383 trip 1 idle segment 12 lasts 28 seconds. Since there are only two observations of such jumps in the whole database, there are not enough data to assess whether they co nstitute a new mode. These observations might indicate that one should pay attention to slow movement during an idle segment. Since these two idle segments show some unusual activities, the modeler will retain them to avoid any bias in the results.

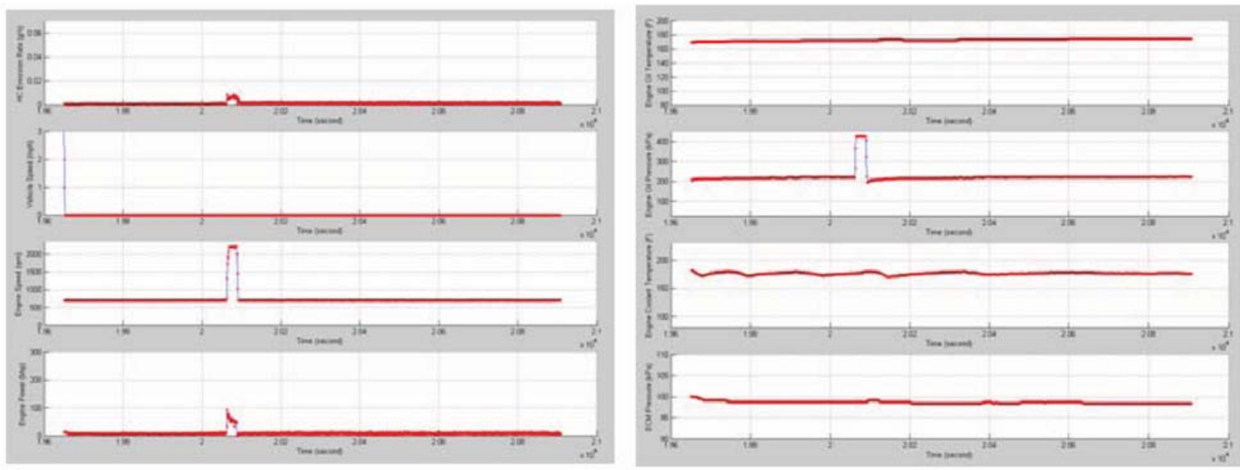


Figure 8-18 Time Series Plot for Bus 383 Trip 1 Idle Segment 12 (1258 Seconds)

8.5 Idle Emission Rates Estimation

Based on definition of “speed $\leq$ 2.5 mph and absolute acceleration $\leq$ 1 mph/s”, about 30% of available data are classified as idle mode. Usually, modelers estimate the idle emission rate by averaging all emission rates in idle mode. Although there are some data points with higher engine power (> 50 bhp) in idle mode, about 90% of data in idle mode exhibit engine power between 0 and 20 bhp. After detailed analysis of all idle segments using time series plots, although some data may be incorrectly classified as the idle mode, no anomalies were noted. To avoid introducing any significant bias, a single idle emission rate is developed for each pollutant. When we treat all data as a whole and put them in the pool, the mean and confidence interval can reflect the distribution of emission rates in real world. Table 8-7 provides idle mode statistical analysis results for NO_x, CO, and HC.

Table 8-7 Idle Mode Statistical Analysis Results for NO_x, CO, and HC

	NO_x	CO	HC
minimum	0.00121	0.00002	0.00001
1st Quartile	0.02201	0.00120	0.00026
mean	0.03342	0.00594	0.00092
median	0.02670	0.00293	0.00051
3rd Quartile	0.03549	0.00554	0.00079
maximum	0.40259	0.48118	0.05232
skewness	4.45050	13.1840	11.6100
Total Number	31064	30764	30250

Due to the non-normality of emission rates, the median value (the value that divides observations into an upper and lower half) and the inter-quartile range (the range of values that includes the middle 50% of the observations) are the most appropriate for describing the distribution. The mean and skewness for the original data are presented in Table 8-8 as well. Although transformation for three pollutants already discussed based on the whole data set in Chapter 6, lambdas chosen by Box-Cox procedure for the whole data set and idle mode are different. Lambdas chosen by Box-Cox procedure for the whole data set are 0.22875 for NO_x, -0.0648 for CO, 0.14631 for HC, while lambdas for idle mode are -0.19619 for NO_x, -0.0625 for CO, 0.002875 for HC. At the same time, using transformation to estimate the mean and construct confidence intervals will create other problems. Therefore the modeler considers bootstrap, another class of general method, to obtain the estimation and construct confidence intervals.

The bootstrap is a procedure that involves choosing random samples with replacement from a data set and analyzing each sample the same way (Li 2004). To obtain the 95% confidence interval, the simple method is to take 2.5% and 97.5% percentile of the β replications T_1, T_2, .., T_β as the lower and upper bounds, respectively. The bootstrap function in this study will resample the emission data 1000 times and compute the mean, 2.5% and 97.5% percentile on each sample. Results are presented in Figure 8-20 and Table 8-8.

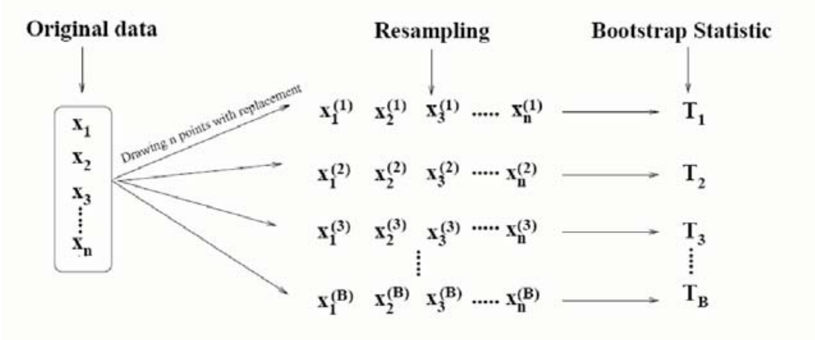


Figure 8-19 Graphical Illustration of Bootstrap (Adopted from Li 2004))

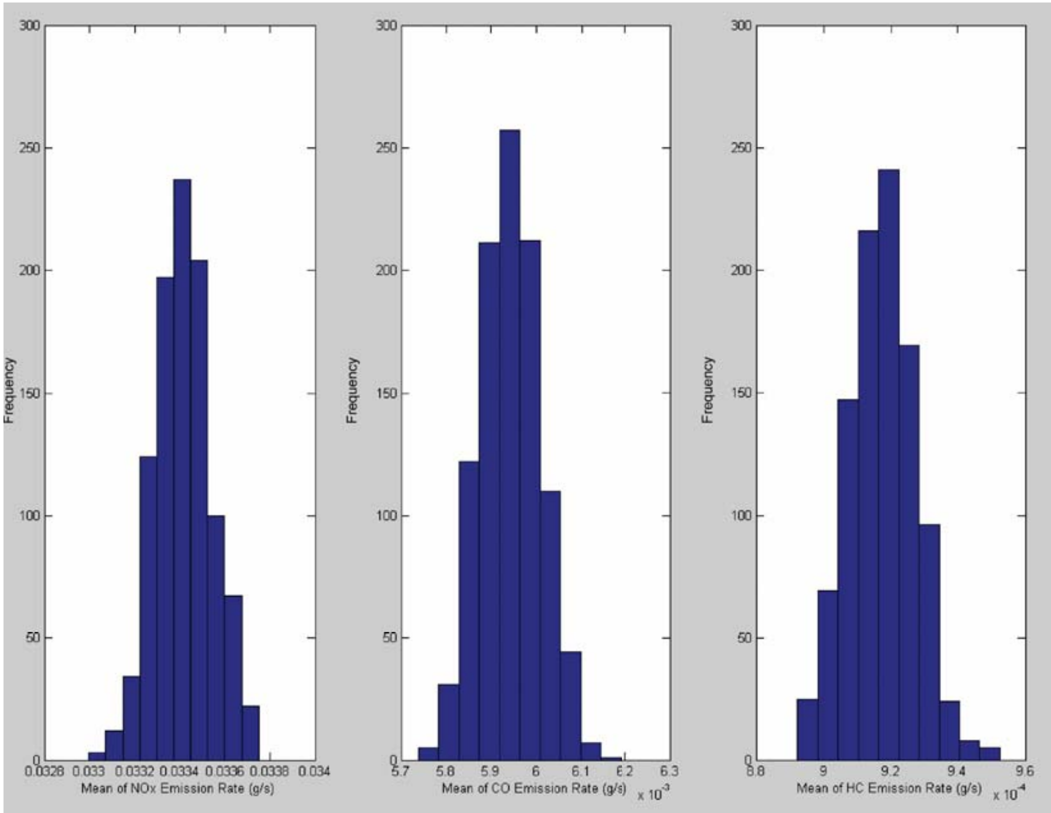


Figure 8-20 Bootstrap Results for Idle Emission Rate Estimation

Table 8-8 Idle Emission Rates Estimation and 95% Confidence Intervals Based on Bootstrap

		Average	**2.5% Percentile**	**97.5% Percentile**
NO_x	Estimation	0.033415	0.010754	0.083266
	Confidence Interval	0.033162	0.010509	0.082279
		0.033669	0.010998	0.084252
CO	Estimation	0.0059439	0.00036116	0.028429
	Confidence Interval	0.0058184	0.00034446	0.028083
		0.0060693	0.00037775	0.028775
HC	Estimation	0.00091777	0.000059167	0.0037260
	Confidence Interval	0.00089742	0.000047572	0.0036412
		0.00093811	0.000070763	0.0038108

Based on table 8-9, the modeler recommends idle emission rates for NO_x as 0.033415 g/s with 95% confidence interval (0.010754, 0.083266), CO as 0.0059439 g/s with 95% confidence interval (0.00036116, 0.028429), HC as 0.00091777 g/s with 95% confidence interval (0.000059167, 0.0037260).

8.6 Conclusions and Further Considerations

In this research, idle mode is defined as "speed ≤ 2.5 mph and absolute acceleration ≤1 mph/s". However the critical value could not be introduced from other research to this research directly. It is more appropriate to test several critical values and obtain the most suitable one instead of testing only one developed from other research.

Inter-bus variability analysis results indicate that bus 372 has the largest mean for NO_x, CO, and HC emissions. Meanwhile, bus 372 has higher minimum (2^{nd}), 1st Quartile (2^{nd}), median (1^{st}), and 3rd Quartile (2^{nd}) engine power by comparison to the other 14 buses. Since test buses and engines are similar in most ways, this difference might be caused by variability of the engines or air conditioner usage. However, the contribution of the air conditioner to engine power in idle mode could not be identified in the database. Future research regarding the role of the air conditioner on engine power and emission rates in idle mode may be able to detect a difference.

Although some trips or some buses have higher mean and standard deviation than others, this kind of variability will decrease when all data in idle mode are treated as a whole. On the other hand, some elevated emissions events may simply reflect real world variability. Without additional evidence, modelers should treat all data as a whole instead of removing outliers and potentially biasing results.

There are two observations of an emissions jump that appears to be unrelated to engine speed, engine power, and engine oil temperature, in a single idle segment. The modeler first assumed that the bus moved too slowly from one location to another location for the GPS/ECM to detect the movement. Other explanations might be an engine computer problem or sensor problem. These two jumps might be evidence to support further research on slow movements during idle segments.

In summary, the modeler recommends idle emission rates for NO_x as 0.033415 g/s with 95% confidence interval (0.010754, 0.083266), CO as 0.0059439 g/s with 95% confidence interval (0.00036116, 0.028429), HC as 0.00091777 g/s with 95% confidence interval (0.000059167, 0.0037260).